HUMAN TRIPLET REPEAT RNA-BINDING PROTEINS: INVOLVEMENT IN mRNA BIOGENESIS AND GENETIC DISEASE

By

JILL WATKINS MILLER

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1998

This work is dedicated to my parents, John and Robbie Watkins and to my husband, David Miller, who love me no matter what.

ACKNOWLEGMENTS

I would like to first thank my high school science teacher, Mr. Lavinghausez, for encouraging me to pursue my ideas and for setting into motion a life-long love of science. I thank Peter Stacpoole, Mary Jo Koroly, and Nancy Denslow for seeing my potential and giving me a place to begin. I would especially like to thank my committee members, John Aris, Henry Baker, Bert Flanegan, Gerry Shaw, and my outside examiner, Jack Griffith, for their interest, help, and guidance. The administrative staff of the Department of Molecular Genetics and Microbiology have been tremendously helpful. I would like to especially thank Bradley Moore and Joyce Conners for making sure that things were done right. I would like to thank all of the members of the Swanson lab for their help and advice and for being such a fun group with which to work. I would especially like to thank Drue DeVore for his tremendous contribution to the project. Most of all, I would like to thank Maury Swanson for his help, enthusiasm, mentorship, and for acting as a fabulous role model. His wisdom, skill, and dedication are something to which I aspire.

TABLE OF CONTENTS

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

HUMAN TRIPLET REPEAT RNA-BINDING PROTEINS: INVOLVEMENT IN mRNA BIOGENESIS AND GENETIC DISEASE

By

Jill Watkins Miller

December, 1998

Chairperson: Dr. Maurice S. Swanson
Major Department: Molecular Genetics and Microbiology

Instability of triplet repeat sequences has recently been recognized as the cause of a number of neurological/neuromuscular diseases. Most of the triplet repeat expansion disorders (TREDs) are dominantly inherited and can be explained as a "gain-of-function" mutation due to an aberrant protein product. Myotonic dystrophy (DM), which also shows autosomal dominant inheritance, results from a (CTG) repeat expansion in the 3' untranslated region (UTR) of a protein kinase gene (DMPK). Several models have been proposed to explain how a $(CTG)_n$ expansion in the 3'-UTR results in an autosomal dominant and variable phenotype. We have focused on the dominant RNA mutation model which proposes that the $(CTG)_n$ expansion acts in a dominant manner at the level of the RNA transcript. The expansion could disrupt the binding of an essential trans-acting factor preventing normal processing of the DMPK transcript. Alternatively, the expansion could act as an abnormal binding site for an RNA-binding protein, thus

sequestering this protein away from its normal function. This project describes the isolation and characterization of two different $(CUG)_n$ RNA-binding activities. The first, hNab50, is a single-stranded RNA-binding protein that may be involved in poly(A) tail length regulation and splice site selection. The second activity, the (CUG) expansion binding (EXP) proteins, are putative double-stranded RNA-binding proteins whose functions are unknown.

The hNab50 protein was originally isolated in a cross-species two-hybrid screen during a study of the yeast hnRNP, Nab2p. Nab2p is important for mRNA 3'end formation and nucleocytoplasmic export of mRNA in yeast. Remarkably, the hNab50 protein is a $(CUG)_8$-binding factor in vitro, and is the first eukaryotic triplet repeat RNA-binding protein to be characterized. The hNab50/CUG-BP protein is a highly conserved heterogeneous nuclear ribonucleoprotein, which is localized predominantly to the nucleus. It binds poly(A)$^+$ RNA in vivo but does not co-purify with the major hnRNP complex suggesting transcript-specific binding activity. While hNab50 is a triplet repeat RNA-binding protein, it also shows transcript-specific crosslinking to DMPK mRNAs in HeLa nuclear extracts but not other $(CUG)_n$-containing transcripts. DMPK transcripts with increased numbers of $(CUG)_n$ repeats show elevated binding of hNab50 in a UV light induced photo-crosslinking assay, although the increase is not proportional to the corresponding increase in repeat length. On the basis of these results, I propose that hNab50 is a polyadenylation factor involved in the regulation of DMPK mRNA 3'-end formation.

In addition to hNab50, another $(CUG)_n$-binding activity was discovered that only photocrosslinks to expanded (CUG) repeats of >20. The expansion binding proteins

(EXP) are specific for $(CUG)_n$ repeats and do not crosslink to $(CAG)_n$ repeats or to the double-stranded transactivation region (TAR) RNA element of Human Immunodeficiency Virus (HIV). Crosslinking of EXP proteins to $(CUG)_n$ repeats of variable size indicates that binding is proportional to repeat size, suggesting that the EXP proteins may bind to an abnormal structure created by the expanded (CUG) repeat. The existence of hNab50 and the EXP proteins as $(CUG)_n$ repeat RNA-binding proteins provides substantial evidence for the dominant RNA mutation model for the pathogenesis of myotonic dystrophy.

INTRODUCTION

Triplet Repeat Expansion Disorders

Human DNA contains an abundance of small repetitive sequences interspersed throughout the genome. Although the polymorphic nature of simple sequence repeats has been known for some time, repetitive sequences have only recently been recognized as the cause of a class of human genetic diseases. Triplet repeat expansion disorders (TREDs) result from expansion of trinucleotide repeat sequences in the context of an affected gene (Reddy and Housman, 1997; Paulson and Fischbeck, 1996). Prior to identification of the molecular defect, these diseases were linked by the genetic phenomenon of anticipation, which is characteristically an increase in disease severity and a decrease in the age of onset with each successive generation. The molecular basis for anticipation was finally elucidated when it was noted that repeat size increased generationally as well. To date there are 12 different disorders that are known to result from the expansion of trinucleotide repeat sequences (Table 1).

These disorders fall into two distinct groups depending on the location of the expanded repeat within the affected gene. For the first group of TREDs (HD, SBMA, DRPLA, MJD, SCA 1, 2, 6, and 7), the mutation occurs in the coding region of the gene resulting in an enlarged polyglutamine stretch and a dominant phenotype. It is believed that the expanded polyglutamines directly cause the neuropathological disease phenotype due to aberrant protein properties that develop at a certain repeat length. For example, a derivative

TABLE 1

Triplet Repeat Expansion Disorders

Disease	Repeat	reference
Huntington's disease (HD)	CAG	The Huntington's Disease Collaborative Research Group, 1993
spinal and bulbar muscular atrophy (SBMA)	CAG	La Spada et al., 1991
dentatorubral-pallidoluysian atrophy (DRPLA)	CAG	Koide et al., 1994; Nagafuchi et al., 1994
Machado-Joseph disease (MJD)	CAG	Kawaguchi et al., 1994
spinal and cerebellar ataxia (SCA) 1,2,6, and 7	CAG	Orr et al., 1993; Imbert et al., 1993; Pulst et al., 1996; Sanpei et al., 1996; Zhuchenko et al., 1997; David et al., 1997
fragile X syndrome	CGG	Kremer et al., 1991; Verkerk et al., 1991
FRAXE mental retardation	CGG	Gecz et al., 1996; Gu et al., 1996
myotonic dystrophy (DM)	CTG	Brook et al., 1992; Mahadevan et al., 1992; Fu et al., 1992
Friedreich's ataxia (FA)	GAA	Campuzano et al., 1996

of the disease gene in MJD containing 78 glutamines causes the formation of nuclear inclusions in a variety of cell types. Neurons are particularly sensitive to the expression of the abnormal protein and a late onset neurodegeneration is seen a *Drosophila* model (Warrick et al., 1998). Additional evidence for an abnormal folding pattern results from aberrant migration of the expressed mutant proteins by gel electrophoresis and the existence of monoclonal antibodies that recognize only expanded polyglutamine tracts (Trottier et al., 1995). Mouse models for HD show that only the first exon need be expressed with the polyglutamine tract under the expression of the HD promoter to produce similar disease symptoms in the mouse (Mangiarini et al., 1996).

The second group of TREDs result from trinucleotide repeat expansion in non-coding regions of the gene and include myotonic dystrophy, Friedriech's ataxia, fragile X syndrome and FRAXE mental retardation. The clinical manifestations of these diseases, and the molecular mechanisms for disease pathogenesis, are somewhat different as compared to the first group. They tend to be more systemic in nature, affecting other organ systems in addition to the neurological or neuromuscular symptoms. For example, fragile X syndrome characteristically results in enlarged ears, head and testicles in addition to mental retardation (Paulson and Fischbeck, 1996). Patients with Freidriech's ataxia usually exhibit cardiac defects and a tendency to develop diabetes in addition to the ataxia that is the identifying feature of this disease (Dürr et al., 1996). Myotonic dystrophy displays a host of multi-systemic defects that will be discussed in detail in the next section. Both fragile X syndrome and Friedriech's ataxia result from the loss of function of the gene product due to disruption of transcription, translation or splicing of the mRNA. Myotonic dystrophy is an unusual disease because the expansion occurs in the 3' untranslated region (UTR) of the myotonic

dystrophy protein kinase (DMPK) gene, yet DM exhibits an autosomal dominant pattern of inheritance.

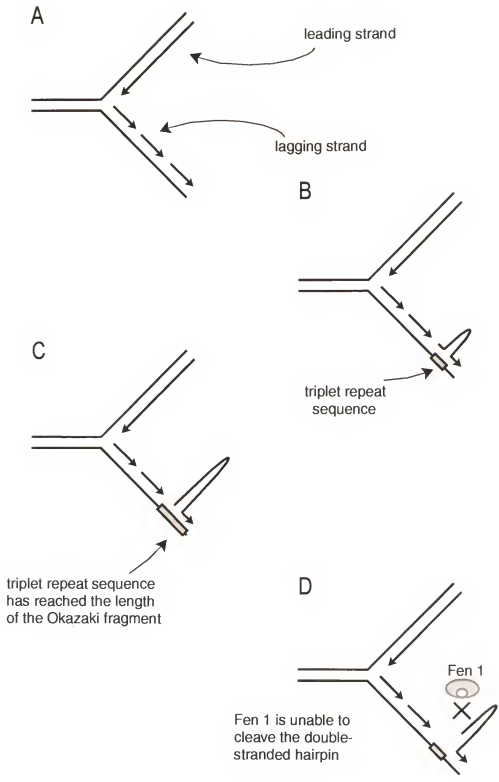
Mechanism of Expansion

Unfortunately, no one truly understands why certain triplet repeat sequences have a particular propensity for expansion, although studies of simple sequence repeats in prokaryotes and eukaryotes suggest certain models. Minisatellite (5 – 100 nucleotides per repeat) and microsatellite (1 – 4 nucleotides per repeat) DNAs are found throughout the human genome and are polymorphic between individuals and populations. Overall, however, the mutation rate for these sequences is not very high (≤15% per gamete) and changes are usually small. Strand slippage during replication, gene conversion, and recombination have all been implicated in the process, and are probably all operative at some level depending on the type of sequence and its location in the genome (Amour et al., 1993). So, what causes the types of mutations that are seen in human disease? Pathologic microsatellite instability can result from two distinct mechanisms (Mitas, 1997). The first is mediated by transacting factors and affects the entire genome. Mutations in proteins, such as those involved in mismatch repair, cause global microsatellite instability resulting in both hereditary and sporadic human cancers (Aaltonen et al., 1993; Peltomaki, 1997). The second mechanism is mediated by the inherent instability of a particular locus. These types of mutations can also result in human cancer (Wooster et al., 1994) and are responsible for the triplet repeat expansion disorders. Although there are examples of other types of repeat expansions (Yu et al., 1997; Lafreniere et al., 1997; Lalioti et al., 1997), trinucleotide repeats appear to be the dominant type of sequence involved in these disorders.

Trinucleotide repeats of certain sequences have been found to form intrastrand hairpin structures that are generally not formed by other types of repeats (Ohshima and Wells, 1997; Smith et al., 1995; Gacy et al., 1995; Mitas et al., 1995; Mitas, 1997). These structures are strongest for triplet repeats that most often cause human disease (CTG*CAG, CGG*CCG, GAA*TTC). Several models have emerged to explain the mechanism by which repeat expansion occurs (See Figure 1). One model is based on studies in *E. coli* using various repeat sizes (Wells, 1996). The ability of CTG repeats to expand or contract depends strongly on the size of the repeat (>30 repeats), its distance from the origin of replication, and the direction of replication. CTG repeats are more likely to expand when they are located on the parental leading strand during replication while the opposite is true when they occur on the parental lagging strand (Kang et al., 1995, Wells 1996). It is thought that the CTG repeat forms a meta-stable hairpin during replication, promoting strand slippage and resynthesis of the repeated sequence (Wells, 1996). This orientation-dependent instability has also been shown to be true in yeast (Freudenreich et al., 1997).

Another hypothesis regarding the propensity for expansion predicts that the instability occurs at the level of Okazaki remodeling (Gordenin et al., 1997; Figure 1D). Once an Okazaki fragment has been generated, the RNA primer must be removed and the newly generated DNA sequence ligated (reviewed in Bambara et al., 1997). Two enzymes important for this remodeling are RNase H1 and Fen 1. Fen 1 is an endo and exonuclease that removes the 5' flap generated during lagging strand synthesis when DNA polymerase encounters an Okazaki fragment . These authors predict that a flap containing CTG repeats would form a hairpin, thus preventing Fen 1 from binding and removing this sequence. This would result in the duplication of this sequence and would explain the propensity for

Figure 1 Models for triplet repeat expansion in the genome. **(A)** Normal DNA replication with the leading and lagging strands indicated. **(B)** Triplet repeat expansion causes a hairpin to form on the lagging strand resulting in strand-slippage and duplication of the repeated sequence. **(C)** The triplet repeat expansion is long enough to span the entire Okazaki fragment. Neither end of the fragment is anchored by unique sequence and several episodes of strand-slippage occur causing a several-fold expansion. **(D)** The hairpin prevents Fen 1 from acting on the flap region of the Okazaki fragment resulting in duplication of the region when the upstream fragment is ligated.


A
leading strand
lagging strand
B
triplet repeat
sequence
C
triplet repeat sequence
has reached the length
of the Okazaki fragment
D
Fen 1
Fen 1 is unable to
cleave the double-
stranded hairpin

expansion over deletion that is seen in triplet repeat disorders. In support of this model, mutations in the yeast homolog of Fen 1 (RAD27) develop duplications and expansions of sequences in their genome (Tishkoff et al., 1997).

Different magnitudes of repeat expansion are seen in human disease. In the majority of cases, where the repeat expansion is located within the coding region of the affected gene, the change in repeat number is relatively small ranging from ~20-100 (Paulson and Fischbeck, 1996). For example, a severely affected Huntington's patient may have 80 repeats while the mildly affected parent has only 40. In the case of expansions that occur in the non-coding regions, in DM or fragile X syndrome, the expansions can be tremendously large. For example, a mildly affected DM patient with 100 repeats could give birth to a severely affected child with over 1000 repeats. The mechanism for these explosive expansions is not understood but one model suggests that once the repeat expansion reaches the full length of an Okazaki fragment, neither end of the repeat is anchored by unique surrounding sequence (Figure 1C). This could lead to strand-slippage at both ends resulting in expansion that occurs on a larger scale (Richards and Sutherland, 1994).

Myotonic Dystrophy

Clinical Manifestations.

Myotonic dystrophy (DM) is classified primarily as a muscular dystrophy due to the progressive nature of the disease that markedly involves the musculoskeletal system (for review see Harper, 1989; Harper and Rudel, 1994). Unlike other muscular dystrophies, DM involves many other organ systems and shows a highly variable phenotype. One of the distinguishing features of DM, and one that sets it apart from other muscular dystrophies, is

the pattern of muscle involvement. Muscles of the face and neck as well as distal musculature of the limbs are involved earliest and most prominently. General weakness of the superficial facial muscles results in a haggard appearance with ptosis (drooping of the eyelids). Anterior neck muscles, sternomastoids, show noticeable wasting. Hollowing of the temples and jaw muscles which can lead to difficulty in chewing as the disease progresses. The palate, tongue and larynx are also affected and can result in difficulty speaking, swallowing and aspiration of material into the bronchi. Intrinsic muscles of the hands and muscles of the wrist and forearm are involved early, showing characteristic wasting. Weight-bearing muscles, such as those in the limb girdle, are usually spared in DM. Microscopic examination of muscle reveals several abnormalities characteristic of DM. Increased centralized nuclei and nuclear chains in muscle fibers, as well as the presence of ringed fibers and sarcoplasmic masses can help distinguish myotonic dystrophy from other types of muscular dystrophy. Other DM specific changes have to do with the ratio of muscle fiber types and size. Type 1 fibers are reduced in size and number and Type 2 fibers show slight hypertrophy. Active degeneration and necrosis are generally not seen and fibrosis is usually only seen late in the disease progression and may be a secondary change. Muscle innervation appears to be normal with no changes in numbers or distribution of acetylcholine receptors.

Myotonia is the hallmark of myotonic dystrophy and can be seen in nearly every patient examined. An exception to this are congenital cases in the first two years of life and some patients with severe wasting, however, it can usually still be detected with the help of electromyography. Myotonia is the inability to relax after a forcible muscle contraction. It is often present long before any other symptoms of the disease are seen and many patients are

unaware that it is an abnormal condition. It can be aggravated by cold weather and is sometimes mistaken for arthritis in older patients.

Myotonic dystrophy is a multi-systemic disease with a characteristic pattern of involvement of several organ systems. Posterior subcapsular cataracts are common, even in the absence of muscle weakness. Defects in smooth muscle reside primarily in the lower pharynx and esophagus leading to swallowing abnormalities and reduced intestinal motility. Cardiac conduction defects are common, particularly heart block and arrhythmias although cardiomyopathy is also seen. Gross pathology of cardiac muscle shows fibrosis of conduction tissue and often fibrosis and fatty infiltration of myocardial muscle as well. Respiratory complications resulting from involvement of the diaphragm and intercostal muscles as well as frequent aspiration of material into the bronchi have a significant impact on patient morbidity and mortality. Pneumonia and cardiac arrhythmias account for the majority of primary causes of death in DM (de Die-Smulders et al., 1998). Many patients present with recurrent chest infections and congenital patients are particularly susceptible to respiratory failure in the first few hours of life. Mild mental deterioration is seen in adults with myotonic dystrophy although overt mental retardation is usually only seen in congenital cases. Testicular atrophy in the majority of male patients results from atrophy, fibrosis, and reduced spermatogenesis in the seminiferous tubules. A high incidence of increased insulin resistance, diabetes mellitus and male baldness is associated with DM.

Genetics

Myotonic Dystrophy is an inherited disease with a frequency of 1:8000 in the population (Harper, 1989). The genetic defect was identified as a (CTG) repeat

expansion in the 3'UTR of a gene located on chromosome 19q13.3 (Brook et al., 1992; Mahadevan et al., 1992; Fu et al., 1992). The normal population has 5 → 37 repeats while affected patients have at least 50 and upwards into the thousands of repeats (Brook et al., 1992). Results of haplotype analysis of nearby markers has led investigators to believe that DM results from a founder mutation (Imbert et al., 1993; Neville et al., 1994). Thus, a subset of the population is more susceptible to a "premutation" that later leads to progressively larger mutations with future generations. The gene codes for a serine-threonine protein kinase termed DMPK for dystrophia myotonica protein kinase (Mahadevan et al., 1993; Shaw et al., 1993). In addition to a kinase domain in the amino-terminus of the protein, a central coiled-coiled domain and a hydrophobic region in the carboxy-terminus were also noted (Brook et al., 1992; Jansen et al., 1992). Homology studies found the DMPK gene to most resemble cAMP-dependent protein kinases, although regulation by cAMP has not been demonstrated (Brook et al., 1992). Additionally, the kinase domain was found to be 45% identical to *Drosophila* gene called *warts* which is thought to act as a tumor suppressor and morphogenic determinant (Justice et al., 1995). Calcifying epitheliomas, which are benign tumors of the hair follicle, have been reported in DM but an increased risk of neoplasia has not been seen clinically (Harper, 1989). The DMPK protein product is most highly expressed in skeletal and cardiac muscle although it has also been found to be expressed in tissues containing smooth muscle and to a small extent in brain (Brook et al., 1992; Jansen et al., 1992; Fu et al., 1993; Jansen et al., 1996). In situ immunohistochemical studies using antibodies against the DMPK protein have demonstrated localization to neuromuscular junctions of skeletal

muscle and intercalated discs of cardiac muscle (Whiting et al., 1995; Maeda et al., 1995; van der Ven et al., 1993). Other investigators have reported a soluble form of the protein (Fu et al., 1993) which may result from alternative splicing at the carboxy-terminus that eliminates the hydrophobic region (Waring et al., 1996). Since several other myotonic syndromes result from genetic defects in sodium and chloride channels (Ptacek et al., 1993), it has been suggested that DMPK regulates the activity of ion channels by phosphorylation. Early studies of red blood cell ghosts and DM muscle biopsy material detected differences in membrane phosphorylation (Roses and Appel, 1973 and 1974). Functional studies *in vitro* have shown that recombinant DMPK possesses serine and threonine phosphorylation activity (Dunn et al., 1994; Timchenko et al., 1995). Other studies have shown alterations in Ca^{++} homeostasis in DM muscle (Jacobs et al., 1990) and in DMPK knockout mice (Benders et al., 1997), but the target of DMPK *in vivo* is unknown. Like other triplet repeat disorders, myotonic dystrophy shows a correlation of disease severity and earlier age of onset with increases in repeat expansion (Redman et al., 1993; Hunter et al., 1992; Jaspert et al., 1995).

Transmission of the repeat expansion from parents to offspring is affected by both the size of the repeat and sex of the parent. Parents with larger expansions show greater instability when transmitting the disease to their offspring (Monckton et al., 1995; Wong et al., 1995). Paternal transmission results in a higher probability of expansion in the offspring, which has been suggested to be due to the increased number of rounds of replication during spermatogenesis (Brunner et al., 1993). However, congenital cases of myotonic dystrophy are almost exclusively transmitted

by the mother (Harper, 1989). This seems to reflect a size limit in sperm, which prevents further expansion past ~1000 repeats (Jansen et al., 1994). Expansion is thought to occur early in embryogenesis since the difference in repeat size between a father's sperm and his child's blood is usually larger than is seen with his own blood (Jansen et al., 1994). Somatic instability also supports a model of expansion during embryogenesis with tissues derived from the same embryonic origin having similar numbers of repeats (Jansen et al., 1994). Both somatic mosaicism in different tissues and mitotic instability over time has been reported by several investigators with muscle containing the largest numbers of repeats (Anvret et al., 1993; Martorell et al., 1995; Wong et al., 1995; Kinoshita et al., 1996).

Models for Disease Pathogenesis

Haploinsufficiency Model

Several models have emerged to explain how an expansion in the 3' UTR of a gene causes an autosomal dominant disease. The DM locus was not found to be imprinted in either human or mouse when a variety of tissues, including skeletal muscle, were tested (Jansen et al., 1993). Haploinsufficiency, or loss of mutant allele expression resulting in only 50% protein production, has been suggested to be the cause (Fu et al., 1993). Many investigators have studied the expression of the DMPK protein product in patient tissues and cell lines. While one investigator reported an increase in protein expression, the general consensus is that repeat expansion has a negative effect on DMPK expression (Fu et al., 1993; Carango et al., 1993; Hofmann-Radvanyi et al., 1993; Novelli et al., 1993). Most DM patients are heterozygotes,

although there are reports of homozygous patients (Martorell et al., 1996; Cobo et al., 1994). These patients exhibit the same abnormalities as heterozygotes suggesting that this disease does not result from simple loss-of-function. Mouse knockouts at the DMPK locus were generated to provide a model for the human disease (Jansen et al., 1996; Reddy et al., 1996). Unfortunately, these animals have not been very informative in terms of the pathophysiology of DM. Single knockout mice, which should best mimic the human disease as heterozygotes, are completely normal. Double knockout mice are normal at birth and only develop mild muscle pathology later in life. Interestingly, a mouse that overexpresses the human gene in a DMPK knockout background exhibited cardiomyopathy although no skeletal muscle defects (Jansen et al., 1996). While partial loss of protein expression of the DMPK gene may be a contributing factor to the disease process, it is probably not the primary cause.

Chromatin Structure Model

The locus surrounding the DMPK gene contains several other genes, and it has been suggested that the triplet repeat expansion not only affects the expression of DMPK but also disrupts local chromatin structure resulting in altered expression of surrounding genes. In support of this model, the CTG expansion has been shown to act as a strong nucleosome positioning element in an *in vitro* reconstitution assay (Wang et al., 1994; Wang and Griffith, 1995). Other studies demonstrate that the repeat expansion alters adjacent chromatin structure in the mutant allele by eliminating a downstream hypersensitive site (Otten and Tapscott, 1995).

Several investigators have examined the expression of the upstream and downstream genes. Gene 59, which is telomeric to DMPK, codes for a protein which contains two regions of WD repeats but no other distinguishable motifs (Shaw et al., 1993; Jansen et al., 1995). The gene products for the mouse homolog of gene 59, DMR-N9, are expressed in brain and testes but are not expressed in muscle (Jansen et al., 1993 and 1995). In addition, Hamshere et al (1997) looked at expression of gene 59 in DM patient cell lines and found its expression to be unaffected. DMAHP, DM associated homeodomain protein, is located just downstream (centromeric) to DMPK (Boucher et al., 1995). While the function of DMAHP is unknown, it is expressed in a wide variety of tissues in the mouse including skeletal muscle, heart testes, brain, and smooth muscle (Heath et al., 1997). Studies of DMAHP in DM have been variable but suggest that there may be an effect of CTG expansion on expression. One investigator saw no difference in expression between DM patients and normal controls (Hamshere et al., 1997). Two other investigators reported a substantial (>2 fold) decrease in the mutant DM-linked DMAHP allele in DM patient derived fibroblasts and myoblasts (Klesert et al., 1997; Thornton et al., 1997). It remains to be seen whether alterations in expression of DMAHP lead to any of the phenotypic symptoms seen in DM.

It has been reported that the CTG repeat occurs within a CpG island that overlaps with the DMPK 3'-end (Shaw et al., 1993; Boucher et al., 1995). A recent study of the methylation pattern of this CpG island was compared between normal patients, adult onset DM patients, and severely affected congenital cases. In the severe congenital cases, hypermethylation was observed in this region as well as loss

of an *in vivo* footprint at a putative Sp1 binding site. Adult onset DM patients and normal controls did not show hypermethylation, suggesting that congenital DM may have a different etiology than adult onset (Steinbach et al., 1998).

RNA Dominant Mutation Model

Several investigators have demonstrated that the mutant allele is transcribed in both cell lines derived from DM patients and from fresh biopsy material (Wang et al., 1995; Krahe et al., 1995; Taneja et al., 1995; Bhagwati et al., 1996; Hamshere et al., 1997; Davis et al., 1997). As a dominantly inherited disease, DM could be caused by a gain-of-function mutation exerted at the RNA level. This would explain why there is a correlation with repeat size and disease severity/age-of-onset and would also explain why DMPK knockout mice have not recapitulated the human disease. An RNA dominant mutation model would also draw a parallel between myotonic dystrophy and the other autosomal dominant TREDs. Instead of an aberrantly structured polyglutamine tract, an aberrantly structured RNA repeat polymer is to blame for disease pathogenesis. Several laboratories have investigated the mutant DMPK transcripts for alterations in RNA metabolism (Wang et al., 1995; Krahe et al., 1995; Morrone et al., 1997; Hamshere et al., 1997; Phillips et al., 1998). One investigator has documented differences in the level of polyadenylation of DMPK transcripts in DM patients (Wang et al., 1995), while another saw changes in the splicing pattern of the troponin T mRNA (Phillips et al., 1998). Three reports have described the accumulation of the mutant allele in the nuclear compartment both by biochemical fractionation and *in situ* hybridization (Taneja et al., 1995; Davis et al.,

1997; Hamshere et al., 1997). In addition, one author reported reduced poly(A)$^+$ RNA levels for insulin receptor mRNA transcripts in DM patient cells, suggesting a "trans" effect of repeat expansion (Morrone et al., 1997). Transgenic mice containing the human DMPK gene with expanded CTG repeats have been created and display intergenerational instability (Gourdon et al., 1997; Monckton et al., 1997b). Although no pathological findings were noted initially, preliminary data on subsequent generations suggest partial DM-like pathology (Monckton et al., 1997a). Finally, proteins that bind specifically to $(CUG)_8$ RNA repeats have been isolated and characterized (Timchenko et al., 1996a and b) and are the subject of this dissertation. It is our hypothesis that $(CUG)_n$ repeat RNA binding proteins are either sequestered on the large repeats or are altered in their function in response to the enlarged transcripts. This sequestration/altered function of (CUG)-binding proteins may be responsible for DM pathogenesis.

DM2 and PROMM

Although 98% of clinically diagnosed cases of DM cases have a trinucleotide expansion in the DMPK gene, a handful of cases have been described which do not have this genetic defect (Mahadevan et al., 1992; Abbruzzese et al., 1996). A disorder that is similar to DM, but with some distinct differences was described a few years ago (Thornton et al., 1995; Ricker et al., 1995). Proximal myotonic myopathy (PROMM) is an autosomal dominant disorder that displays many of the multi-systemic defects seen in DM, such as cataracts, but muscle involvement has a different distribution and quality. Proximal muscles, particularly of thigh, limb girdle

and arms are involved while facial and hand muscles are spared. While myotonia and muscle weakness are features of PROMM, atrophy is not seen in this disorder. Interestingly, anticipation may be associated with PROMM. Although only a few families have been studied, worsening of the condition in offspring was seen in over half of those studied (Ricker et al., 1995). The genetic locus for PROMM has not been isolated, however, mutations in DMPK or in loci of other myotonic diseases (myotonia congenita and paramyotonia) were ruled out. Recently, a family has been described with several affected members having a clinical phenotype identical to DM, but no triplet repeat expansion (Ranum et al., 1998). The genetic defect in this form of myotonic dystrophy, termed DM2 by the authors, maps to a 10 cM region of chromosome 3q. This region is distinct from loci known to cause other forms of myotonia. Isolation of the mutation of both PROMM and DM2 could answer many questions about involvement of the DMPK protein product, and the repeat expansion in DM disease.

Nucleic Acid Triplet Repeat Structures

Many studies have been devoted to understanding the unusual structures that CTG triplet repeats form in DNA (Gacy et al., 1995; Mitas et al., 1995; Smith et al., 1995; Petruska et al., 1996; Mariappan et al., 1996). One group of investigators studied single-stranded CTG oligonucleotides of 25 repeats using NMR spectroscopy (Gacy et al., 1995). These authors found that the DNA molecules form stable intra-strand hairpins in solution and that the mispaired T-T bases form hydrogen bonds and are highly stacked within the stem. Another study that utilized electrophoretic mobility, chemical and enzymatic probing

methods also found strong evidence for intra-strand hairpin structures formed by CTG repeats (Mitas et al., 1995).

(CUG) repeat RNAs have not been well studied, however, one study has demonstrated that CUG repeats possess many of the same properties as their DNA relatives (Napeirala and Krzyzosiak, 1997). These authors studied the DMPK 3'-UTR region containing CUG repeats of 5, 11, 21, or 49 by Pb probing, as well as S1 and T1 nuclease digestion. They found that CUG repeats of 5 were completely single-stranded under all conditions tested. RNAs with 11 CUG repeats displayed transient and unstable hairpin structures at low temperatures, and were completely single-stranded at higher temperatures. RNAs of 21 and 49 repeats formed stable double-stranded hairpin structures that could be maintained at moderate temperatures in a length dependent manner. The $(CUG)_{49}$ hairpins were stable up to 75° C. Cleavage patterns along with computer modeling studies led the authors to conclude that CUG repeats > 21 repeat units form double-stranded hairpin structures with 4 – 7 bases in the single-stranded loop region. Moreover, the stability of these structures increases as repeat length increases. The inability of the probing agents to cleave the mismatched U-U base pairs in the $(CUG)_{21}$ and $(CUG)_{49}$ stem region suggests that the backbone of these RNAs is much more rigid than smaller repeats. Other investigators have found U-U mismatched base pairs in small RNA duplexes to be unexpectedly stable. In one study, X-ray diffraction was used to study the U-U base pairs in duplexes of the dodecamer GGACUUUGGUCC (Baeyens et al., 1995). These authors found that although there were four non-Watson-Crick base pairs (two of which are U-U), there was no gross distortion of the alpha double helix and the U pairs were hydrogen bonded. Another group used NMR to study a U-U mismatch in a conserved hairpin loop of the large ribosomal

subunit and found it also to be hydrogen bonded and stacked (Wang et al., 1996). This particular U-U mismatch is conserved between prokaryotes and eukaryotes suggesting that this type of mispair may have a biological function.

Nuclear Pre-mRNA Metabolism

Protein/RNA Recognition

The importance of RNA structures and the specific interactions that proteins have with these RNAs is basic to the function of biological systems. RNA-binding proteins are fundamental to the post-transcriptional control of gene expression. Proteins interact with RNAs in a multitude of ways, often altering RNA structure in the process. There are many recurrent themes in terms of the protein domains that directly bind RNA, however, new and different variations on these themes are being discovered at a rapid pace.

The best characterized consensus RNA-binding domain is the RNP consensus sequence (RNP-CS) RNA binding domain (CS-RBD) (Dreyfuss et al., 1988; Bandziulis et al., 1989). This motif consists of two highly conserved consensus sequences, RNP1 and RNP2, in the context of a ~ 90 amino acid stretch. The crystal structure for the amino terminal RBD of U1A protein reveals a binding surface consisting of four anti-parallel β-pleated sheets and two α-helices (β1-α1-β2-β3-α2-β4) (Nagai et al., 1990). Conserved amino acid residues on the surface of the beta sheet "platform" have been found to interact directly with the RNA bases (Allain et al., 1996) although residues outside of the beta sheets are important for binding specificity. The CS-RBD is capable of recognizing several different types of RNA

substrates, from single-stranded RNAs, hairpin loops and internal loops. Its recognition can be highly sequence specific, with changes in only two nucleotides in U1 snRNA determining U1A or U2B'' binding (Scherly et al., 1990). There are over 300 identified proteins that contain known or putative CS-RBDs. The majority of hnRNPs contain at least one of these motifs, and snRNA binding proteins, ribosomal RNA-binding proteins, and many other proteins involved in pre-mRNA or rRNA processing also contain CS-RBDs.

The RGG box consists of an approximately 20 amino acid stretch containing the repeated tripeptide of arginine-glycine-glycine interspersed with aromatic residues. A variety of hnRNP proteins, as well as nucleolar proteins, contain this motif, often in conjunction with other RNA-binding motifs. While crystallographic studies have not yet been reported, other structural studies suggest that the RGG box forms a 'β-spiral' and is able to disrupt RNA secondary structures by unstacking the bases (Ghisolfi et al., 1992; Kiledjian et al., 1994).

The KH domain was originally described in the hnRNP K protein (Siomi et al., 1993). It consists of a 60 amino-acid stretch containing several highly conserved residues. NMR spectroscopy studies reveal the structure of the KH domain to be three anti-parallel β-sheet packed against three α helixes on one face (Musco et al., 1996). FMR1, the protein responsible for fragile X syndrome contains two KH motifs which are important for RNA binding. FMR1, along with two structurally related interacting proteins FXR1 and FXR2 also containing KH motifs, have been found associated with the 60S ribosomal subunit (Siomi et al., 1996).

A double-stranded RNA binding domain (ds-RBD) has been recently described and NMR studies have revealed that it forms three anti-parallel β-pleated sheets flanked by two α-helices (Kharrat et al., 1995). This structural organization is reminiscent of the CS-RBD described above. Double-stranded RNA-binding proteins are found in a variety of organisms and have a variety of binding sites. The *Drosophila* Staufen protein was one of the first ds-RNA binding proteins characterized and binds to a double-stranded stem loop structure in the 3' UTR of the *bicoid* mRNA. Staufen anchors the *bicoid* mRNA at the anterior of the cell insuring proper localization for pattern formation in the *Drosophila* embryo (St Johnson, 1995). TRBP (TAR RNA-binding protein) contains two ds-RBDs and was isolated as a result of its ability to bind to the HIV TAR element (Gatignol et al., 1991). Mutagenesis studies have revealed that only the second ds-RBD is needed for TAR binding and that this domain recognizes a GC rich stem within the TAR stem-loop structure (Gatignol et al., 1991,1993).

Arginine-rich motifs as well as zinc-finger and zinc-knuckle motifs have been found to bind RNA. The HIV Rev protein requires several arginine residues to make base-specific contacts in the major groove of its double-stranded target, the RRE (Rev Response Element), in the HIV genome (Battiste et al., 1996). A particularly interesting example of a zinc-finger RNA binding protein is TFIIIA which binds both the 5S ribosomal RNA gene as well as 5S ribosomal RNA. It acts as a transcription factor in the expression of the 5S rRNA in addition to being a major component of a 7S RNP particle in *Xenopus* oocytes (Moore, 1996).

Heterogeneous Nuclear Ribonucleoproteins: Structure and Function

The most abundant class of nuclear RNA binding proteins in metazoans is the heterogeneous nuclear ribonucleoproteins (hnRNPs). They are defined as proteins whose stable and primary binding site is hnRNA (Dreyfuss et al., 1988). Heterogeneous nuclear RNAs (hnRNA) are products of RNA polymerase II (pol II) which are heterogeneous in size and are localized to the nucleus. HnRNPs associate with nascent pol II transcripts during transcription to form hnRNP complexes (Amero et al., 1992; Matunis et al., 1993). Many different methods have been employed to biochemically analyze hnRNP complexes. Initially, cosedimentation with hnRNA on sucrose gradients was used to identify several of the more abundant components, the A, B, and C proteins, in HeLa cells (Beyer et al., 1977). Later, ultra-violet (UV) light induced photocrosslinking in vivo followed by isolation on oligo (dT) cellulose, was used to isolate hnRNPs (Mayrand et al., 1981; Choi and Dreyfuss, 1984). UV irradiation and purification of hnRNP proteins allowed for the isolation of hnRNPs that are directly bound to nuclear poly(A)$^+$ RNAs *in vivo*. This method corroborated data that was obtained by sucrose gradient sedimentation and also allowed for the generation of monoclonal antibodies against specific hnRNPs. These antibodies have allowed for the direct immunopurification of hnRNP complexes, and has enabled the identification of other hnRNP proteins that do not efficiently crosslink or were lost during sedimentation (Pinol-Roma et al., 1988). The major hnRNP complex isolated from HeLa cells contains over 20 different types of protein (>50 total) most of which have been purified and characterized (Dreyfuss et al., 1993). All bind directly to RNA and are designated hnRNP A through U.

Analysis of the primary sequence of hnRNPs has revealed them to be extremely diverse in structure (see Dreyfuss et al., 1993 for review). Each possesses a unique combination of RNA binding motifs (as discussed above) and auxiliary domains, giving them what has been called a "modular" structure. Auxiliary domains are thought to mediate protein-protein interactions, thus possibly recruiting processing factors or affecting the localization the RNA *in vivo*. These auxiliary domains have been compared to activation domains of transcription factors in that they often contain stretches of certain amino acids or types of amino acids (such as glutamine or proline).

With such diversity in structure, it would follow logically that hnRNP proteins would also show diversity in substrate binding specificity. Early studies using ribohomopolymer substrates established that members of the hnRNP complex possess affinities for different RNA substrates (Swanson and Dreyfuss, 1988a and 1988b). In addition, studies in *Drosophila* revealed that the stoichiometry of hnRNP proteins on nascent RNA pol II transcripts varies with the type of transcript (Matunis et al., 1993). Finally, selection of short RNAs by hnRNP A1 in vitro reveals high affinity binding sites for specific sequences (Burd and Dreyfuss, 1994). Thus, hnRNPs are a diverse group of RNA-binding proteins that possess both general and specific RNA-binding properties.

The functions of hnRNPs is not well understood. Their sheer abundance suggests that they might have a structural role as packaging proteins for hnRNA (Beyer et al., 1977). Although this may be one function of hnRNPs, their structural diversity and binding specificities alone would argue against this being their only

function. In addition, hnRNPs have been found to be involved in a variety of post-transcriptional processes including splicing, polyadenylation, and mRNA export (Swanson, 1995).

Although there are ~20 abundant hnRNPs that make up the major hnRNP complex in cells, many other factors involved in different stages of RNA metabolism could also be classified as hnRNPs. Some of the proteins characterized as RS proteins, splicing enhancers, or factors involved in alternative polyadenylation site choice could also be classified as hnRNPs although they are less abundant than classical hnRNPs. They are predominately nuclear in their subcellular distribution and their primary binding site is hnRNA.

hnRNPs in *Saccharomyces cerevisiae*

HnRNPs have been isolated and characterized from many different organisms and have recently been identified in yeast (Anderson et al., 1993). Four nuclear polyadenylated RNA-binding (Nab) proteins have been characterized and they are all essential for viability (Anderson et al., 1993; Wilson et al., 1994; Krecic, 1998). Yeast hnRNPs contain many of the same structural motifs as those in metazoans although their arrangement within the protein is somewhat different.

Metazoan hnRNPs have been extensively studied over the last twenty years in terms of their structure, binding properties, and subcellular localization. Functionally, these proteins have been extremely difficult to understand due to the lack of an in vitro system by which to characterize them. In an attempt to better understand the function of hnRNPs in pre-mRNA processing and maturation, hnRNPs from the yeast

Saccharomyces cerevisiae have been isolated and characterized (Anderson et al., 1993; Anderson et al., 1994, Wilson et al., 1994). *S. cerevisiae* offers the possibility of studying hnRNPs genetically as well as biochemically and may provide more information in terms of the specific processes in which these proteins are involved. Four yeast hnRNPs have been described and are referred to as the nuclear polyadenylated RNA-binding proteins, or Nab proteins (Anderson et al., 1993; Anderson et al., 1994, Wilson et al., 1994; Krecic, 1998).

Capping

Immediately following initiation of transcription by RNA pol II, nascent pre-mRNA transcripts begin to undergo a variety of processing events mediated by a multitude of factors. They must be modified at both 5' and 3'ends and often have internal sequences removed to produce a mature mRNA molecule. One of the first events in this process is capping of the 5' end of the transcript (Saldítt-Georgieff et al., 1980). Capping occurs co-transcriptionally and serves many purposes including the protection of the 5' end from nucleases, involvement in pre-mRNA processing, export, and translation (reviewed in Lewis and Izaurralde, 1997). The cap structure consists of an inverted 7-methyl guanosine that is joined by a 5'-5' triphosphate bond. The capping enzyme removes the terminal phosphate of the nascent RNA molecule and subsequently transfers a GMP residue to the 5'-diphosphate on the RNA molecule. Finally, RNA (guanine-7)-methyltransferase methylates the cap resulting in a $m^7G(5')ppp(5')N$ cap or m^7cap (Shatkin, 1985; Mizumoto and Kaziro, 1987). The major cap binding proteins in the nucleus are CBP80 and CBP20, which form the cap-binding complex (CBC) (Shatkin, 1985). The CBC has been found to play a role in pre-

mRNA splicing of the cap-proximal intron (Izaurralde et al., 1994; Lewis et al., 1996). This coupling of capping to pre-mRNA splicing is consistent with the exon definition model (Robberson et al., 1990) which states that coordination of splicing is facilitated by factor-factor interaction across exons (rather than introns). In the case of the cap, there is no 3' splice site to define the exon, thus the cap acts to facilitate this process (Lewis et al., 1996; Lewis and Izaurralde, 1997). A recent study has demonstrated that the CBC influences 3'-end processing as well (Flaherty et al., 1997). Immunodepletion of CBC proteins inhibited the cleavage of an L3 substrate in HeLa nuclear extracts and addition of recombinant protein restored activity. Capping of mRNA transcripts also affects their export from the nucleus and is important for translation as will be discussed in later sections.

Pre-mRNA Splicing

Pre-mRNA splicing is the removal of intronic or intervening sequences from pre-mRNA transcripts and the joining of exons (for review see Krämer, 1996). Under normal circumstances, removal of introns is required for proper mRNA maturation and its subsequent export to the cytoplasm for translation into protein. Splicing occurs in both lower and higher eukaryotes, but higher eukaryotes usually contain more introns per transcript. Cis-acting sequences that are needed for efficient splicing include a 5' splice site, 3' splice site, branch site, and usually a polypyrimidine tract just upstream of the 3' splice site. These sequences are highly degenerate in mammalian cells, adding complexity to the reaction.

The basic splicing reaction begins with the 2' hydroxyl group of the adenosine at the branch point reacting with the 3',5'-phosphodiester bond at the 5'

splice site by nucleophilic attack. The 3' hydroxyl that is generated at the 5' splice site attacks the phosphodiester bond of the 3' splice site resulting in the joining of the two exonic regions and release of the lariat formed by the first reaction (Green, 1991; Krämer, 1996). In mammalian cells, constitutive splicing involves the coordinated efforts of five major small nuclear ribonucleoprotein particles (snRNPs; designated U1, U2, U4/U6, and U5) and several associated trans-acting factors. Formation of the spliceosomal complex, which catalyzes the splicing reaction, begins with the recognition of 5' and 3' splice sites by U1 snRNP and U2AF (U2 snRNP auxiliary factor). This initial reaction requires the efforts of several serine/arginine rich proteins (SR proteins) that bind to the pre-mRNA and recruit U1A and U2AF. U2 snRNP associates with the branch site and also requires additional factors, including U2AF, to promote stable base pairing. The final step in spliceosome formation prior to catalysis involves the incorporation of U5 snRNP and U4/U6 snRNP. U6 snRNA, which is extensively base paired with U4, must be rearranged to base pair with U2 snRNA, to form a catalytically competent structure (Krämer, 1996). Once the splicing reaction is complete, the snRNPs are released with the lariat structure.

Many protein factors are required for efficient constitutive and alternative pre-mRNA splicing. One large family of proteins are the arginine/serine-rich splicing factors (RS proteins; Fu, 1995). Many of these proteins contain RNA-binding domains (usually of the CS-RBD type) and all contain a domain rich in serine and arginine, often as a dipeptide repeat. These proteins are highly conserved and are essential for constitutive splicing. Models for RS protein function suggest that some of these proteins bind to the pre-mRNA prior to spliceosome formation and act to

recruit the U snRNPs to the proper sites. Others are thought to associate with the U snRNPs directly and escort them to the forming spliceosome. It has been proposed that the functions of RS proteins are redundant depending on the pre-mRNA involved. Some pre-mRNAs require the presence of specific RS proteins for proper splicing while others will splice in the presence of any number of RS proteins (Fu, 1995).

HnRNPs have also been found to play roles in splicing. HnRNP A1 shows specific binding to sequences that resemble both 5' and 3' splice sites (Burd and Dreyfuss, 1994). It has also been found that the concentration of hnRNP A1 relative to another splicing factor ASF/SF2 is important in splice site choice (Mayeda and Krainer, 1992). The polypyrimidine tract binding protein (PTB/hnRNP I) was isolated as a factor that binds to the polypyrimidine tract and has been implicated in alternative pre-mRNA splicing as discussed below (Ghetti et al., 1992).

Alternative Pre-mRNA Splicing

Alternative splicing is a powerful strategy for higher eukaryotes to exponentially expand the possibilities of a limited genome. Pre-mRNAs can be alternatively spliced in many different ways. Exons can be included or excluded by alternative 3' or 5' splice site choice, or by skipping entire groups of exons (McKeown, 1992; Cooper and Mattox, 1997). Cooperation of cis-acting sequences and trans-acting factors determine when, where, and how a pre-mRNA is spliced. Cis-acting sequences can be located within exons or introns and can act as either positive regulatory sequences (enhancers) or negative regulatory sequences

(repressors). For example, in cardiac troponin T (cTNT), exon five is included in embryonic cardiac muscle but is excluded in adult. Muscle-specific splicing enhancers are located both upstream and downstream of exon five. Transient transfection experiments have demonstrated that exon inclusion requires trans-acting factors that are only found in embryonic tissue and not in adult (Ryan and Cooper, 1996).

Neurons are notorious for using alternative splicing in post-transcriptional gene regulation. One particular group of proteins, termed the Elav-like proteins because of their structural similarity to the *Drosophila* Elav protein (embryonic lethal abnormal visual system), has figured prominently in neuronal specific expression and alternative splicing. The *elav* gene is expressed only in neurons and is important in the development of the *Drosophila* nervous system (Campos et al., 1985; Robinow and White, 1988). Recently it has been shown that loss of the Elav protein corresponds with the alternative splicing of a neural specific form of the protein, Neuroglian (Koushika et al., 1996). Whether Elav works as an enhancer or repressor for Neuroglian is still not known.

The human Hu RNA-binding proteins (HuB, HuC, and HuD), structural homologs of *Drosophila elav*, also are neuronal in expression. These proteins were originally isolated using antisera from patients suffering from autoimmune neurodegenerative disorders. These proteins are not only neural specific, but show regional-, cell- and developmental-specific expression patterns in the mouse (Okano and Darnell, 1997). While a direct role in alternative splicing has not been

demonstrated, overexpression of HuD in chick neuroblasts accelerated their differentiation (Wakamatsu and Weston, 1997).

Characterization of the sequences that direct splice site choice and exon inclusion/exclusion are under intense study, particularly in the central nervous system. Both cis-acting repressor and enhancer sequences work to promote the selection of neuron-specific exons. For example, exon skipping in the γ2 subunit of the $GABA_A$ receptor utilizes several cis-acting repressor sites, located around the branch site of the exon, to prevent the inclusion of the exon in rat cerebellum. These repressor sites have been found to bind a ubiquitously expressed hnRNP, PTB/hnRNP I (Ashiya and Grabowski, 1997). In neurons, binding of PTB results in the formation of a repressor complex and exclusion of the exon. Removal of PTB by the addition of competitor RNAs results in the formation of a spiceosomal complex at the site and inclusion of the exon. Although PTB is ubiquitously expressed, several different isoforms of the protein have been characterized and display tissue specific expression patterns (Ashiya and Grabowski, 1997; Grabowski, 1998). Differential binding of the PTB isoforms to the repressor sequences is still under investigation.

These data support the idea that there are general splicing factors and transcript/tissue specific splicing factors. The general splicing factors, such as the RS proteins, can affect splice site selection or enhance splicing for some pre-mRNAs based on their relative concentration (A1 and ASF/SF2 as discussed above). However, these factors are ubiquitously expressed and harbor the ability to provide redundant functions in general pre-mRNA splicing. Specific splicing factors are often expressed in a tissue specific manner, have specific or preferential binding sites

on a particular pre-mRNA or subset of pre-mRNAs. These factors may be specific isoforms of proteins that are more widely expressed, such as suggested for PTB, or they may be primarily found in one specific cell type. The regulation of alternative pre-mRNA splicing is reminiscent of the elaborate mechanisms used in transcriptional control. The words 'enhancer' and 'repressor' were initially the language of transcription, and functionally many parallels can be drawn. All cells (with a few exceptions) have the same DNA, and thus the same cis-acting sequences. It is the differential expression of trans-acting factors, enhancers and repressors, that often determine the fate of the cell.

It is also becoming obvious that tissue/cell type specific splicing is important in human disease. For example, EAAT2 is a glutamate transporter that is implicated in amyotropic lateral sclerosis (ALS). Aberrant splicing of the EAAT2 pre-mRNA has been found in patients with ALS, although no defect in the primary genetic sequence has been isolated. It has been speculated that the defect lies in a neuron-specific splicing factor resulting in improper splicing of EAAT2 pre-mRNA (Grabowski, 1998; Lin et al., 1998).

3' End Formation

All mRNAs in eukaryotes are polyadenylated with the exception of histone mRNAs in metazoans. Pre-mRNA 3' end formation involves two processes, cleavage at a specific site at the 3' end of the RNA and addition of a poly(A) tail (reviewed in Wahle and Keller, 1996; Colgan and Manley, 1997). The length of the poly(A) tail varies in different organisms from 250-300 (A) residues in vertebrates to

70-90 residues in yeast. Functionally, the poly(A) tail has been shown to positively influence the translatability and stability of an mRNA (Jackson and Standart, 1990; Ross, 1995). The poly(A) tail is the first target in the deadenylation-dependent decay pathway of mRNA degradation and its length can affect its half-life. In addition, it is thought that the poly(A) tail, or at least the act of its formation, functions in the export of the mRNA out of the nucleus (Eckner et al., 1991).

Pre-mRNAs contain cis-acting sequences that are required or promote efficient cleavage and polyadenylation. The polyadenylation signal AAUAAA is one of the most conserved of the sequence elements in vertebrates and is located 10-30 nucleotides upstream of the cleavage site. This sequence is highly invariant in vertebrates with only 10% of mRNAs differing (usually by only one nucleotide). A second cis-acting sequence is a highly degenerate GU or U rich region that is located usually within 30 nucleotides downstream of the cleavage site in metazoans. Yeast possess a much more degenerate polyadenylation signal. The first element is located approximately 20 nucleotides upstream of the cleavage site and is known as the positioning element. It most likely corresponds to the AAUAAA in higher eukaryotes although it can be highly variable in yeast. The second, called an efficiency element, is often AU-rich and is located upstream of the positioning element (Wahle and Kühn, 1997). Interestingly, although the majority of mammalian genes contain a GU or U-rich sequence downstream of the polyadenylation signal, there are examples of genes that utilize an upstream element instead (Brackenridge et al., 1997; Moreira et al., 1998). The existence of upstream elements is presumably due to the close proximity of a downstream gene. This argument has also been used

to explain the differences seen between lower and higher eukaryotic polyadenylation elements.

Mammalian polyadenylation *in vitro* requires five separate factors as depicted in Table 2 (Keller and Minvielle-Sebastia, 1997; Wahle and Kühn, 1997). In vitro assays involving biochemical fractions has allowed the processes of cleavage and poly(A) addition to be dissected and the necessary components identified. Cleavage requires the participation of four factors, CPSF, CstF, CF I, and CF II. Cleavage and polyadenylation specificity factor (CPSF) binds to the conserved polyadenylation element AAUAAA, although this binding is weak in the absence of the cleavage stimulation factor (CstF), which interacts with CPSF and stabilizes the complex. All subunits of CstF bind to RNA, although the 64 kD subunit is specific for the G/U rich sequences found dowstream of the polyadenylation signal. Cleavage factor I (CF I) is required for cleavage and all of its subunit polypeptides possess RNA-binding activity although its role is not understood. Cleavage factor II (CF II) is also required for cleavage but little is known about its components or function. The cleavage reaction requires ATP but it is unknown which component of the cleavage machinery acts as the endonuclease (Colgan and Manley, 1997).

Poly(A) polymerase (PAP) catalyzes the synthesis of poly(A) addition although it will nonspecifically add A residues to any substrate in the absence of the other factors. Polyadenylation begins as a distributive process until ~ 10 (A) residues are added, at which time it switches to processive synthesis (Sheets and Wickens, 1989). PAP is an inefficient enzyme on its own and is stimulated by CPSF and an additional factor, poly(A) binding protein II (PAB II) (Wahle, 1995). In an in vitro

TABLE 2

Protein Factors Required for Mammalian Polyadenylation

Factor	Process	Polypeptide (kD)
CPSF	cleavage and polyadenylation	160 100 73 30
CstF	cleavage	77 64 50
CF I	cleavage	68 59 25
CF II	cleavage	unknown
PAP	cleavage and polyadenylation	77/82
PAB II	poly(A) tail length control	33

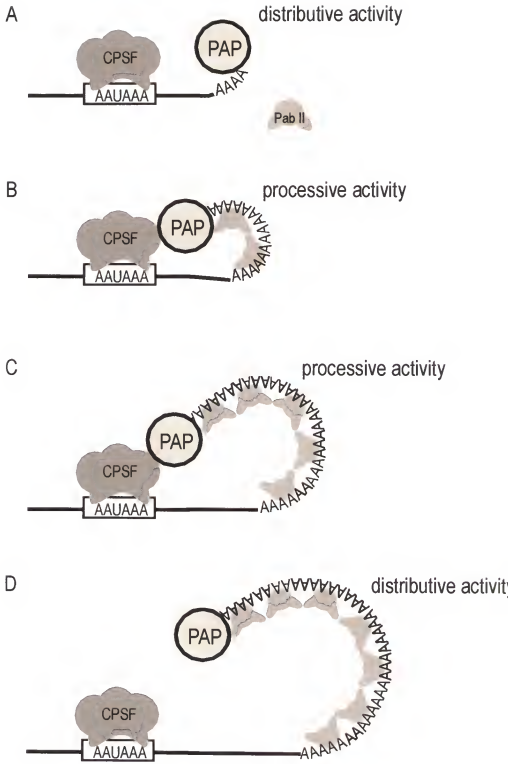
data derived from Keller and Minvielle-Sebastia, 1997

assay system using a pre-cleaved substrate, CPSF and PAB II promoted the processive activity of PAP. Regardless of the starting length of the poly(A) tail, rapid synthesis was seen up to ~250 residues. After this length was reached, further elongation was 300 times slower. The current model for poly(A) tail length regulation suggests that tail length is measured by the number of PAB II molecules bound (see Figure 2). Once the tail reaches 250 residues, the CPSF-PAB II complex dissociates (Wahle and Kühn, 1997).

Polyadenylation is emerging as another means for gene regulation, analogous to alternative splicing (Proudfoot, 1996; Edwalds-Gilbert and Milcarek, 1997). There are many examples of genes that have two or more tandem polyadenylation sites. These sites can differ in the strength of the cis-acting sequences which would affect the efficiency of cleavage and polyadenylation depending on the concentration of certain polyadenylation factors within that particular cell (Edwalds-Gilbert and Milcarek, 1997). Another means of regulation involves the coupling of alternative splicing and polyadenylation site selection.

Immunoglobulin genes are a well studied example of alternative polyadenylation. B cells must switch from producing a membrane-bound form of IgM to a secreted form as they mature into plasma cells. Inclusion of an upstream exon containing a weak polyadenylation signal results in the secreted form while alternative splicing that joins a downstream exon with a strong site results in the membrane bound form. Surprisingly, the switch from membrane to secreted forms correlates with an increase in the level of the 64 kD subunit of CstF (Takagaki et al., 1996). The efficiency of the site may determine the site choice since the upstream

Figure 2. Model for poly(A) tail length regulation in mammalian cells. **(A)** Depicted is the 3' end of a newly cleaved pre-mRNA. Cleavage and polyadenylation specificity factor (CPSF) is bound to the polyadenylation element (AAUAAA). Poly(A) polymerase (PAP) has added the first few A residues in a distributive fashion since PAB II is not yet able to bind. **(B)** PAP has switched to processive synthesis with the binding of PAB II and the interaction with CPSF. **(C)** Processive synthesis continues as long as PAP can interact with both PAB II and CPSF. **(D)** Once the poly(A) tail has reached 250 residues, PAP is unable to contact CPSF due to steric hindrance created by the binding of PAB II to the elongating tail.


A
distributive activity
CPSF
PAP
AAUAAA
AAAA
Pab II
B
processive activity
CPSF
PAP
AAUAAA
C
processive activity
CPSF
PAP
AAUAAA
D
distributive activity
PAP
CPSF
AAUAAA

weak site may only be efficiently utilized in the presence of elevated levels of polyadenylation factors (Proudfoot, 1996; Edwalds-Gilbert and Milcarek, 1997).

Nuclear polyadenylation is the primary method utilized for producing poly(A) tails on mRNA molecules. However, specialized mechanisms have evolved to control the level of polyadenylation of particular transcripts within the cytoplasm (Richter, 1996). Cytoplasmic polyadenylation occurs in eggs and embryos of a variety of species and involves its own set of regulatory sequences and proteins. Silencing or activation of maternal mRNAs at particular times during development is accomplished by either the specific removal or addition of a poly(A) tail on particular transcripts. Cytoplasmic polyadenylation elements (CPE) have been defined for several mRNAs. These elements are U rich and usually occur in the 3'-UTR at a position 10-50 base pairs upstream of the polyadenylation signal (AAUAAA). Both the CPE and the polyadenylation signal are necessary for cytoplasmic polyadenylation (Richter, 1996). Likewise, deadenylation elements have also been described that target an mRNA to lose its poly(A) tail at a precise developmental time. For example, the *Xenopus* mRNA c-mos contains an embryonic deadenylation element (EDEN) in its 3'-UTR and is deadenylated shortly after fertilization. Binding of a 53/55 kD protein, designated EDEN-BP, to this element correlates with rapid deadenylation of this mRNA (Paillard et al., 1997).

Nuclear mRNA Export

Fully processed mRNA molecules must be exported from the nucleus so that they can be translated on ribosomes in the cytoplasm. These RNAs are exported out

of the nucleus through the nuclear pore complex (NPC), an enormous structure that spans the nuclear membrane. The nuclear pore complex is made up of at least 100 different proteins arranged in an 8-fold symmetrical cylindrical structure. Small molecules can diffuse freely across through the nuclear pore, however, larger molecules (>60 kD) must use an energy-dependent process (Davis, 1995). Export of RNA from the nucleus has been difficult to study due to the lack of a good *in vitro* system. Injection studies in *Xenopus* oocytes, genetic studies in *Saccharomyces cerevisiae*, as well as direct visualization of Balbiani ring transcripts in *Chironomus tentans*, have provided most of the information to date (Nakielny et al., 1997).

Electron microscopy has been extremely helpful in both determining the structure of the NPC and in understanding how macromolecules traverse it. Early studies showed that gold particles coated with different types of RNA molecules could be seen passing through the NPC (Dworetzky et al., 1988). Studies of the large Balbiani ring (BR) transcripts in *Chironomus tentans* have been revealing in terms of current models of mRNA export. This 35-40 kb mRNA is large enough to visualize as a discrete entity as it is transcribed, processed, and exported to the cytoplasm (Visa et al., 1996; Daneholt, 1997). During transcription, the BR pre-mRNA is bound by hnRNP proteins and other processing factors. Once the transcript has been spliced, it proceeds to form a tightly packed RNP particle of 50 nm in size. Once this particle reaches the NPC, it is reoriented, unfolded and is passed through the NPC as a linear "ribbon" with the 5' end in the lead. The BR mRNA is engaged by ribosomes concurrently with its emergence into the cytoplasm (Visa et al., 1996; Alzhanova-Ericsson et al., 1996). Nuclear mRNA export in *Chironomus tentans* is an

exciting model system that supports the current thinking about hnRNP particles and the possible roles of hnRNPs in mRNA export.

Much more is known about protein import than export because an *in vitro* assay system exists and has been tremendously useful in dissecting this process. Two protein import pathways have been defined, each with its own factors that must be cycled back and forth between the nucleus and the cytoplasm. The importin/karyopherin protein import system was described initially (reviewed in Nigg 1997; Gorlich and Mattaj, 1996). Importin α/karyopherin α associates with a basic nuclear localization signal (NLS) located on the protein to be imported (or to an adaptor) and then binds importin β/karyopherin β to facilitate docking at the NPC (see Table 3). Translocation requires GTP hydrolysis by the Ran GTPase. The second system employs one protein, transportin, which recognizes the M9 sequence on hnRNP A1 and facilitates its import into the nucleus (Pollard et al., 1996). While the transportin system seems to parallel the importin/karyopherin system, the emerging consensus import sequence (M9) recognized by transportin is quite different (see Table 3; Siomi et al., 1998). The M9 sequence was first defined in the carboxy terminus of hnRNP A1 and acts as the signal for both import and export (Siomi and Dreyfuss 1995; Michael et al., 1995). Interestingly, transportin interacts with several hnRNPs including hnRNP A1, hnRNP F and hnRNP D0 (Siomi et al., 1997; Siomi et al., 1998). The yeast homolog of transportin, Kap104, interacts with two yeast hnRNPs, Nab2p and Nab4p (Aitchison et al., 1996; Siomi et al., 1998).

Export of mRNA from the nucleus is poorly understood, although mRNA is almost certainly exported in association with RNPs as suggested by studies in *C.*

Table 3

Nuclear Localization Signals	
SV 40 T antigen[a]	P**KKKRK**V
nucleoplasmin[a]	KRPAATKKAGQA**KKKK**
Nuclear Import/Export Signals	
hnRNP A1[d]	N**QSS**NF**GPM**K**GGNFGGRSSG** **PYGGG**GQYFAKPRNQ**GGY**
HRP36[d]	QGGGSGGWNQ**QGGS**GG**GPWNN** Q**GGGNGGWNGGGGGGGYGGG**
HRP40[d]	GYGYGGGFEGNGYGGGG**GGNM** **GGGRGGPRGGG**GPK**GGGFNGG**
Nab2p[d]	APVDNSQRFT**QRGGGAVGKNRRGG** **RGGNRGGRNNNSTRFNPLAKALG**
Nuclear Export Signals	
PKI[a]	**LALKLAGLDI**
Rev[a]	**LPPLERLTL**
Mex67p[c]	**LELLNKLHL**
Gle1p[b]	**LPLGKLTL**

Depicted are the consensus sequences for selected nuclear import and export signals. Bolded letters indicate conserved amino-acids. The M9 sequence of hnRNP A1 is both an import and an export signal. These sequences have not been established as export signals for HRP26, HRP40, or Nab2p.
[a] Gerace, 1995; [b] Murphy and Wente, 1996; [c] Segref et al, 1997;
[d] Siomi et al., 1998

tentans (Daneholt, 1997). Recent advances in the understanding of protein export from the nucleus have contributed to ideas about the mechanism of mRNA export. Leucine-rich nuclear export sequences were defined several years ago due to their interaction with cellular adaptors that facilitate their export (see Table 3; Fritz and Green, 1996; Stutz et al., 1995; Bogerd et al., 1995). Initial protein/RNA export studies centered around viral systems which have contributed much to our understanding. The human immunodeficiency virus (HIV) Rev protein facilitates the export of unspliced HIV RNAs from the nucleus (Cullen and Malim, 1991). Normally, unspliced pre-mRNAs are retained in the nucleus (Nakielny et al., 1997) and thus the virus had to evolve a method to circumvent this problem. In addition to producing RNAs with inherently inefficient splicing signals to slow down the splicing process, the virus also utilizes the Rev protein to quickly export the unspliced RNAs from the nucleus (Cullen and Malim, 1991). Rev contains a leucine-rich NES (Table 3) and binds to a cellular protein hRIP/Rab (Rev-interacting protein or Rev activation domain binding protein; Stutz et al., 1995; Bogerd et al., 1995). The hRIP/Rab protein contains a phenylalanine/glycine repeated sequence (FG) that is found in many nuclear pore proteins and is transiently associated with the NPC. By interacting with NPC associated proteins, Rev and its RNA cargo are efficiently exported from the nucleus.

Recently, an RNA export signal was defined in the Mason-Pfizer monkey virus (MPMV) (Ernst et al., 1997). This element is called a constitutive transport element (CTE) and acts to export both viral RNAs and cellular intron containing RNAs without the use of any viral (such as Rev) adaptors (Ernst et al., 1997). In

other words, this RNA signal relies on a cellular factor to bind to the CTE sequence and export the RNA into the cytoplasm. Addition of CTE RNA blocks export of mRNA, suggesting sequestration of factors needed for mRNA export (Pasquinelli et al., 1997).

Another export factor has been recently identified in both yeast and metazoans. Exportin 1 (Crm1p or Xpo1p) interacts with the leucine-rich NESs directly and functions in their export from the nucleus (Stade et al., 1997; Fornerod et al., 1997). There is also evidence that this interaction is dependent on the presence of Ran (Fornerod et al., 1997). In addition, yeast *crm1* mutants accumulate poly(A)$^+$ RNA in the nucleus suggesting a direct role in mRNA export in yeast (Stade et al., 1997).

Overall Goal of Research Project

The overall goal of this project was to gain an understanding of the functional roles of hnRNPs in mRNA metabolism and human disease. Specifically, my interest focused on the role that triplet repeat RNA binding proteins might have in the pathogenesis of myotonic dystrophy. It is my hypothesis that the defect in myotonic dystrophy involves the disruption of the nuclear metabolism of both the DMPK mRNA and possibly other mRNAs. This research project centered on the isolation and characterization of triplet repeat RNA-binding proteins and their role in DM disease.

The initial part of this study focused on identifying two types of CUG repeat RNA-binding proteins. Characterization of the first protein, hNab50, revealed it to be

an hnRNP. I showed that hNab50 not only binds CUG repeats but associates with DMPK RNA. I hypothesized that hNab50 is involved in poly(A) tail length regulation and provide supporting evidence. The hNab50 protein associates with Nab2p, a yeast hnRNP, in the two-hybrid system and a discussion of the significance of this interaction is included. Next I focused on the binding of hNab50 with mutant DMPK RNAs containing various numbers of expansions. The study of DMPK mutant transcripts led to the identification of the expansion binding proteins or EXP proteins, a different type of CUG repeat RNA-binding protein. Characterization of these proteins and the implications to DM disease are discussed.

MATERIALS AND METHODS

Media and Cell Culture

All animal cells, unless otherwise indicated, were grown at 37° celsius (C) in the presence of 5% CO_2 in a NAPCO model 5430 incubator. Human S3 and JW36 HeLa cells were cultivated in DMEM (BRL) supplemented with 5% calf serum (BRL) and 100 units/mL penicillin/streptomycin [(P/S) BRL]. HeLa cells, grown in spinner flasks, were grown at 37° C using Joklik modified MEM (BRL) with 5% calf serum and 100 units/mL of P/S. Rabbit RK13, human Hep2, and mouse NIH 3T3 cells were grown in DMEM supplemented with 10% fetal bovine serum (Hyclone) and 100 units/mL P/S. Normal and DM patient lymphoblast cell lines (GM03696C, GM03756, GM03928, and GM03986) were obtained from Coriell Cell Repositories. The normal lymphoblast line (HH) was obtained from the Tissue Culture Core Laboratory at Baylor College of Medicine. Lymphoblasts were grown in RPMI 1640 (BRL) supplemented with 20% FBS and 100 units/mL P/S. Normal and DM myoblasts (SW, KB, and Cab) were kindly provided by Dr. Luba Timchenko (Baylor College of Medicine) and were cultivated in Ham's F-10 media (Hyclone) supplemented with 15% FBS, 5% calf supplemented defined serum (Hyclone) and 100 units/mL P/S. The A549 cells were grown in MEM (BRL) supplemented with 10 mL of glutamine (200 mM), 10 mL of sodium pyruvate (100 mM), 1% non-essential amino acids (BRL) and 8% FBS. Chicken CEF cells were grown in DMEM supplemented with 10% tryptose phosphate broth (29.5 g/L), 5% calf

serum, 5% chicken serum (heat inactivated), 1% gentamycin (10 mg/mL stock). Xenopus XL1 cells were grown at 30° C in DMEM supplemented with 8% FCS and 100 units/mL P/S.

Unless otherwise specified, all plasmid amplifications were carried out in *E. coli* strain DH5α (*supE*44Δ*lac* U169 [Ø*lacZ*ΔM15] *hsdR*17 *recA*1 *endA*1 *gyrA*96I *thi*-1 *relA*1) or DH10B [F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) *lacZ*ΔM15 *lac*ΔX74 *deo*R *rec*A1 *ara*Δ139 Δ(*ara,leu*) 7697 *gal*U *gal*K λ–*rps*L *end*A1 nupG]. Yeast strains HF7c [*MATa ura*3-52 *his*3-Δ200 *lys*2-801 *ade*2-101 *trp*1-Δ901 *leu*2-3 –112 *gal*4-Δ542 *gal*80-Δ538 LYS2::GAL1-HIS3 URA3::(GAL4 17-mers)3-CYC1-lacZ] and BJ926 (*MATa/MATα prb1-1122/prb1-1122 prc1-407/prc1-407 pep4-3/pep4-3 can1/can1 gal2/gal2 his1/HIS1 TRP1/trp1*) were either grown in YPD (1% Bacto-yeast extract, 2% Bacto-peptone, 2% dextrose) or synthetic dextrose (SD) media (0.67% Baco-yeast nitrogen base without amino acids, 2% dextrose) supplemented with the following amino acids: 20 mg/L adenine, 30 mg/L leucine, 20 mg/L uracil, 20 mg/L tryptophan, 30 mg/L lysine and 20 mg/L histidine.

^{35}S-Labeling of Cells

HeLa S3 cells were grown to subconfluence in DMEM supplemented with 5% calf serum and 1% P/S. Cells were washed twice with sterile phosphate buffered saline [(PBS) 0.14 M NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.7 mM KH_2PO_4, pH 7.4] and then incubated in labeling media [DMEM without methionine or cysteine, 5% calf serum, 1% P/S plus 20 μCi/mL of ^{35}S-methionine (Dupont), or TRAN^{35}S-methionine (Dupont)] for 12 – 18 hours.

Preparation of Total Cellular Proteins and Nuclear Extracts

Two methods were utilized to prepare total cellular proteins. For analyzing proteins by SDS-PAGE and immunoblotting, 1 mL of SDS-PAGE loading buffer (125 mM Tris, pH 6.8, 20% glycerol, 4% SDS, and 100mg bromphenol blue) was added to one 10 cm plate of confluent cells, cells were scraped with a rubber policeman and transferred to a microcentrifuge tube. Samples were sonicated three times for 10 seconds each using a Vibracell sonicator (Sonics Materials, Danbury Connecticut). Proteins were heated to 100°C for 3 minutes, spun at 12,000 x *g* for 5 minutes and loaded directly onto an SDS-PAGE gel. The second method was used for preparing total cellular proteins for use in *in vitro* crosslinking/label transfer experiments. One confluent plate of HeLa cells was washed with PBS and then scraped using a rubber policeman into 2 mL of PBS. Cells were pelleted at 750 x g for 4 minutes and resuspended in 300 µL of Buffer E (50 mM Tris-HCl, pH 8.0, 0.1 mM EDTA, 1 mM DTT, 12.5 mM $MgCl_2$, 20% glycerol, 0.1 M KCl, 1% Triton X-100, 1 µg/mL leupeptin/pepstatin). After incubating on ice for 5 minutes, samples were spun at 4300 x *g* for 4 minutes and the supernatant was collected and frozen at -80°C (Timchenko et al, 1995).

Preparation of HeLa nuclear extracts was carried out essentially as described by Dignam et al., (1983). Four liters of HeLa cells, grown to a density of 4×10^5 cells/mL in spinner flasks, were pelleted at 1000 x *g*, washed with PBS and repelleted. Cells were resuspended in 5 pellet volumes of Buffer A [10 mM Hepes, pH 7.9, 1.5 mM $MgCl_2$, 10 mM KCl, 0.5 mM dithiothreotal (DTT)] and incubated on ice for 10 minutes. Cells were pelleted at 1000 x *g* for 10 minutes at 4°C, resuspended in 2 pellet volumes buffer A and

lysed by 10 strokes of a Dounce homogenizer with a type B pestle. The cell homogenate was spun at 1000 x *g* to pellet nuclei and, after removal of supernatant, the pellet was respun at 25,000 x *g* for 20 minutes at 4°C to pack pellet. Nuclei were resuspended in 3 mL of buffer C (20 mM Hepes, pH 7.9, 25% glycerol v/v, 0.42 M NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT) for every 1 X 10^9 cells and homogenized by 10 strokes of a Dounce homogenizer with type B pestle on ice. The homogenate was slowly stirred at 0° C for 30 minutes and then spun for 30 minutes at 25,000 x *g* at 4°C. The supernatant was dialyzed against at least 1000 volumes of buffer D (20 mM Hepes, pH 7.9, 20% glycerol v/v, 0.1 M KCl, 0.2 mM EDTA, 0.5 mM PMSF, 0.5mM DTT) for 5 hours at 4°C. The dialysate was spun for 20 minutes at 25,000 x *g* at 4°C, quick frozen in liquid nitrogen, and stored at -80°C.

Cell Transformation and Plasmid Rescue

Plasmids were introduced into *E. coli* by electroporation using a gene pulser (BioRad, Richmond, CA) according to manufacturer's instructions. Small scale yeast transformations were carried out essentially as described (Ito et al., 1983) with minor modifications. Overnight cultures grown in YPD were diluted to an OD_{600} of 0.25 in fresh prewarmed media and cells were allowed to grow to an OD_{600} of 1.0. Ten OD units of cells were pelleted at 1,500 x *g* and washed with 10 mL of sterile water. Cells were resuspended in 100 µL of ice cold LiOAc/TE solution (10 mM Tris, pH 7.5, 1 mM EDTA, 10 mM lithium acetate, pH 7.5). To this solution, 25 µg of calf thymus DNA (boiled for 10 minutes and quick-cooled on ice) and 3-5 µg of plasmid DNA were added and mixed on ice. Finally, 600 µL of PEG/LiOAc solution (40% polyethelene $glycol_{3350}$,

10 mM Tris, pH 7.5, 1mM EDTA, 10 mM lithium acetate, pH 7.5) were added to cells, mixed well on ice, and then cells were heat shocked for 15 minutes at 42° C. Cells were then mixed with 400 μL 1 M sorbitol, pelleted, and washed again with 1M sorbitol. Cells were resuspended in 1 mL of 1 M sorbitol and 1/10 of the transformation was spread on agar plates containing the appropriate drop-out media. Cells were incubated at 30° C for 48 – 72 hours.

Large scale yeast transformations for two-hybrid screening were carried out as outlined by the manufacturer (Clontech). Briefly, 200 mL of HF7c (containing pGBT9-NAB2) cells were grown to an OD_{600} of 0.5 in SD-Trp dropout media. This culture was used to inoculate 1 L of prewarmed YPD supplemented with adenine (20 mg/mL). Cells were harvested by centrifugation at 1,500 x *g* at an OD_{600} of 1.0, washed with 200 mL of TE (10 mM Tris, pH 7.5, 1 mM EDTA) and repelleted. Cells were resuspended in 20 mL of TE/LiOAc solution and incubated for 10 minutes at 30°C. Ten mg of calf thymus DNA and 500 μg of purified HeLa S3 Matchmaker cDNA library DNA (Clontech.) were added and cells were incubated for 10 minutes at 30°C. To this mixture, 140 mL of PEG/LiOAC solution was added and cells were incubated for 30 minutes at 30°C. Finally, 17.6 mL dimethylsulfoxide (DMSO) was added and cells were heat shocked for 10 minutes at 42°C with occasional swirling, and then rapidly cooled in ice water. Cells were pelleted, washed once with 50 mL of TE, and resuspended in 500 mL of prewarmed YPD. The cells were allowed to recover at 30°C with shaking for 1 hour, pelleted, washed twice with 50 mL of TE, and then resuspended in 5mL of TE. Cells (100μL) were plated onto SD-Trp-Leu-His plates and incubated for 72 hours at 30°C.

Plasmid rescue was performed as described previously (Strathern and Higgins, 1991). DNA was further purified using a Gene Clean kit (Bio101, Vista, CA) according to the manufacturer's instructions.

Isolation of PINs

PIN (proteins that interact with Nab2p) were isolated using the a commercial yeast two-hybrid interaction system (Clontech). The yeast strain HF7c was transformed with pNAB2.GBT9 by the small-scale method described above and fusion protein expression confirmed by immunoblot analysis using anti-Nab2p mAb 3F2 (Anderson et al, 1993). Cells expressing the Gal4pDBD-Nab2p fusion protein were subsequently transformed with a HeLa cell cDNA library cloned into the pGADGH plasmid as described above (large scale). Cells were selected on SD-Leu-Trp-His plates, and clones were initially tested for β-galactosidase activity using a plate assay following a protocol provided by the manufacturer. Cells that were the most blue by the plate assay were subsequently tested for β-galactosidase activity using a quantitative liquid assay (see below). All positives were tested for self-activation by removing pNAB2.GBT9 and re-testing by the plate assay. Plasmids were recovered by plasmid rescue, amplified in *E. coli,* and the human DNA inserts sequenced.

Quantitative β-Galactosidase Liquid Assay

Positive clones were grown in 10 mL of selective media (SD-Leu-Trp-His) to an A_{260} of 0.5. Cells were harvested at 1,500 x *g* at 4° C and immediately frozen at -80° C until ready to assay. To assay, cells were washed in 1 mL of ice-cold Z-buffer (0.1 M

$NaHPO_4$, pH 7.0, 0.1 M KCl, 10 mM $MgSO_4$, 1 mM DTT) and resuspended in 100 μL of Z-buffer. An equal volume of glass beads was added and cells were vortexed four times for 30 seconds, with a 10 second incubation on ice in between. Samples were spun at 10,000 X g for 10 minutes at 4° C and the supernatant was transferred to a fresh tube. An additional 50 μL of Z-buffer was added to the remaining beads and the process repeated. The supernatants were combined and mixed. A protein assay was performed on 10 μL (in duplicate) of the sample using Bradford's reagent (Biorad) according to the manufacturer's protocol. For the β-galactosidase assay, 100 μL of sample was placed in a 13 mm test tube and 900 μL of Z-buffer were added and mixed. The reaction was begun by adding 200 μL of ONPG (4 mg/mL in 0.1 mM $NaHPO_4$, pH 7.0) and transferring immediately to a water bath at 30° C. Samples were incubated from 10 minutes to 2 hours depending on the rate of color development. Samples that developed in less than 10 minutes were diluted and re-assayed. The reaction was stopped by adding 500 μL of 1 M $NaCO_3$ and the absorbance at 420 nm was determined. Results were expressed as specific activity: A_{420}/(0.0045) (mg/mL protein) (volume in mL) (time in minutes).

cDNA Library Screening

To obtain a full-length clone of hNab50/Pin22, an EcoR I/Kpn I restriction fragment of the two-hybrid clone was uniformly labeled with [^{32}P]-dCTP using a random-primed labeling kit (BRL) according to the manufacturer's instructions. An osteosarcoma (Lambda Zap; Stratagene, La Jolla, CA) cDNA library (prepared by Dr. Maurice Swanson), a human fetal brain library (kindly provided by Dr. Thomas Yang),

and a human HeLa S3 cDNA library (Lambda Zap; Stratagene) were screened. Colony filters were hybridized in 50% formamide, 5XSSC (1XSSC=0.15 M NaCl, 0.015 M sodium citrate, pH 7.0), 0.2% SDS, 5X Denhardt's solution [1% Ficoll, 1% polyvinylpyrrolidine, 1% BSA (fraction V)], 100 μg/mL salmon sperm DNA at 42°C. Filters were washed twice in 2X SSC, 0.1% SDS at room temperature for 15 minutes with shaking. A third wash was carried out in 0.5X SSC, 0.1% SDS at 65°C for 30 minutes. Filters were exposed to film overnight and positive plaques were picked using a pasteur pipet and stored in 1 mL of SM (50 mM Tris-Hcl, pH 7.5, 10 mM $MgSO_4$, 100 mM NaCl, 0.01% gelatin) with 20μL chloroform added at 4°C. To obtain plasmid from the lambda phage, *in vivo* excision was carried out using *E. coli* SOLR cells (Stratagene, La Jolla, CA) according to the manufacturer's protocol. Three cDNA clones encoding the putative full-length hNab50 protein were isolated from both human osteosarcoma and the HeLa cell libraries. The osteosarcoma full length clone (hNab50.20) was sequenced extensively on both strands using gene specific primers. The Genbank accession number for hNab50 is U63289. Isolation of the RPL14 cDNA was performed by Miltiadis Paliouris and is described in his Masters Thesis (Paliouris, 1998).

Expression screening for EXP proteins was carried out using a CUG54 uniformly labeled RNA probe (see in vitro transcription and plasmid constructs). Filters were prepared from a HeLa cDNA library (Lambda Zap; Stratagene) as described (Snyder et al., 1987) and were either denatured in 6M guanidine-HCl in binding buffer (30 mM HEPES, pH 7.9, 50 mM KCl, 1.4 mM $MgCl_2$,0.1 mM EDTA, 0.4 mM DTT) or were placed directly in binding buffer for 1 hour prior to probing. For filters that were denatured in 6 M guanidine-HCL (in binding buffer), renaturation of proteins was

accomplished by subsequent 15 minutes incubations in decreasing concentrations of guanidine-HCl (6 M, 3 M, 1.5 M, 0.75 M, and 0.375M) followed by two washes in binding buffer and pre-incubation in binding buffer for 1 hour. Filters were then incubated with the RNA probe (50-100 pmol/mL) for 1-2 hours in binding buffer. Filters were washed 5 times in binding buffer and positives were visualized by autoradiography.

Purification of CUG-BP and Bandshift Analysis

CUG-BP purification and bandshift analysis was performed in the Timchenko lab and is described elsewhere (Timchenko et al, 1996a and b)

Monoclonal Antibody Preparation

For the preparation of anti-hNab50 polyclonal antisera, BALB/c mice were injected with a hNab50-maltose-binding-protein (hNab50-MBP) fusion protein which was prepared by expression of the pMAL50.1 plasmid in *Escherichia coli* TB1 cells followed by amylose resin affinity chromatography (New England Biolabs) according to the manufacturer's protocol. The pMAL50.1 plasmid was constructed by cloning a partial hNab50 cDNA clone (encoding amino acids 44-482) behind the *malE* gene. Antisera were tested by immunoblot analysis using both purified hNab50-MBP protein as well as HeLa whole cell lysates. The mAb 3B1 was prepared by the University Florida Interdisciplinary Center for Biotechnology Research (ICBR) Hybridoma Laboratory. Hybridoma supernatants were screened by immunoblotting and cellular immunofluorescence.

SDS-PAGE and Immunoblot Analysis

For immunoblot analysis of hNab50, proteins were fractionated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) on a 12.5 % acrylamide separation gel (Laemmli, 1970). Proteins were then transferred to a nitrocellulose membrane (Schleicher and Schuell) using a semi-dry electroblotter (Bio-Rad Laboratories, Hercules, CA) as suggested by the manufacturer. For immunoblotting, the membrane was blocked with blotting milk [10% non-fat dry milk and 0.5% Nonidet P-40 (NP40) in PBS] for 1 hour at room temperature, and subsequently incubated with mAb 3B1 (1:500) or mAb 3F2 (1:500) diluted in blotting milk. After washing three times with PBS + 0.5% NP40, membranes were incubated for 30 minutes with a sheep anti-mouse secondary antibody conjugated with horseradish peroxidase and washed several times with PBS. Proteins were detected by ECL (Amersham Corp.) according to the manufacturer's instructions.

Indirect Cellular Immunofluorescence

Indirect cellular immunofluorescence was performed essentially as described (Choi and Dreyfuss, 1984; Wilson et al., 1994) by growing HeLa cells, normal and DM myoblasts directly on sterile 10-well HTC Blue slides (Cel-Line Associates, Newfield, NJ) followed by exposure to 2% formaldehyde in PBS for 30 minutes. After washing three times in PBS, slides were incubated for 3 minutes in cold acetone followed by three washes in PBS. Proteins were detected by incubating cells for 1 hour with a 1:500 dilution of 3B1 or 1D8 [specific for hnRNP M proteins (Datar et al, 1993)] diluted in 3% BSA/PBS at room temperature. After washing three times in PBS, slides were incubated with a fluorescein-conjuged goat anti-mouse IgG1 (Cappel) at a dilution of (1:10) in 3%

BSA/PBS for 30 minutes at room temperature. For visualizing DNA, slides were washed three times in PBS and then incubated with 0.5 μg/mL of 4'6-diamidino-2-phenylindole [(DAPI) Sigma, St. Louis, MO] in PBS. Slides were mounted by applying mounting media [1 mg/mL p-phenylene dianine (Sigma) and 90% glycerol] to each well, covered with a glass coverslip and sealed with clear nail polish. Slides were visualized using a Nikon Optiphot-2 microscope equipped with a 100X fluorescence/differential interference contrast (DIC) objective.

Cell Fractionation and Immunoprecipitation of hnRNP Complexes

HeLa S3 cells were labeled with Tran ^{35}S-methionine as described above. The culture medium was aspirated and the cells were washed twice with cold PBS. The cells were scraped using a rubber policeman into 1 mL of cold buffer A [10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2.5 mM $MgCl_2$, 0.5% aprotinin (Sigma), 2 μg of pepstatin A per mL, 2 μg of leupeptin per mL, and 0.5% Triton X-100] per 10 cm plate and homogenized by four passages through a 25 gauge needle. The nuclei were pelleted at 3000 x *g*, resuspended in 0.5 mL of cold buffer A, and sonicated twice for 5 seconds each using the microtip of a sonicator (Vibracell; sonics materials) on ice. The sonicate was layered on a 30% sucrose cushion (prepared in buffer A) and centrifuged at 4000 x g for 15 minutes. The supernatant, defined as the nucleoplasm, was used for subsequent immunopurifications.

Immunopurification of hnRNP complexes was carried out using MAb 4F4 or 3B1. Antibodies (2.5 μL 4F4 or 6 μL 3B1) were attached directly to 25μL protein A-Sepharose for one hour in RSB-100 [10 mM Tris-HCl, pH 7.4, 100 mM NaCl, 2.5 mM

$MgCl_2$] plus 1% Triton X-100 at 4° C. The beads were washed 3 times with RSB-100/1% Triton X-100 and incubated with nucleoplasm at 4° C for 15 minutes with gentle rocking. The beads were washed four times by resuspension in 1 mL of RSB-100/1%Triton X-100 and the bound material was eluted from the Sepharose beads with 50 µL of SDS-Page loading buffer. Samples were fractionated by SDS-PAGE and visualized by fluorography. To determine if hNab50 was part of the hnRNP complex, the complex was first isolated using MAb 4F4 as just described except that the final precipitate was eluted off of the beads by boiling for 3 minutes in 1% SDS. Samples were then diluted in PBS containing 1mM EDTA, 1% Triton X-100, 0.5% deoxycholic acid, 0.1% SDS, 0.5% aprotinin and subjected to a second round of immunoprecipitation using the 3B1 antibody.

Nucleic Acid Methods

Small-scale purification of plasmid DNA was carried out by the alkaline lysis procedure (Sambrook et al., 1989). Large-scale plasmid preparations were performed using a commercial plasmid isolation kit (Wizard Midi-prep; Promega, Madison, WI) according to the manufacturer's instructions. Restriction digests, ligations, and agarose gel electrophoresis were carried as described by Sambrook, et al (1989). DNA sequencing was carried out using a sequencing kit (Sequenase$_{TM}$, Amersham Corp.) using either T7, SP6 or gene specific primers as outlined by the manufacturer.

Poly $(A)^+$ RNA isolation from HeLa S3 cells was performed on 15 – 20 sub-confluent plates of cells by washing plates twice with RSB-100 [10Mm Tris-HCl, pH 7.4, 100 mM NaCl, 2.5 mM $MgCl_2$] and then lysing cells directly in 2.5 mL lysis buffer

[RSB-100 containing 1% SDS, 0.5% 2-mercaptoethanol (βME), 10 mM vanadyl ribonucleoside complex, and 350 μg/mL proteinase K (Boehringer Mannheim, Indianapolis, IN)] per plate. Twenty plates of sub-confluent lymphoblasts were pelleted at 750 x *g* for 5 minutes, washed once with RSB-100 and then resuspended in 1-2 mL of RSB-100 to which 50 mL of lysis buffer was added. Lysed cells were homogenized by 10 strokes with a type A pestle in a glass Dounce homogenizer and incubated at 42° C for 30 minutes. After addition of EDTA to a final concentration of 10 mM, cells were incubated for an additional 15 minutes at 42° C. The lysate was then incubated at 65° C for 10 minutes followed by rapid cooling on ice. LiCl (10 M) was added to a final concentration of 0.5 M and lysates were spun at 4,500 x *g* for 10 minutes at 4° C to pellet unwanted cell debris. Supernatants were mixed with 0.2 g of oligo(dT)-cellulose (Gibco/BRL) in binding buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.5% SDS, 0.5 M LiCl) and incubated for 30 minutes at room temperature or at 4° C with nutation. Poly(A)$^+$ RNA was eluted from the oligo(dT)-cellulose by column chromatography using elution buffer (10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.05% SDS). Poly(A)$^+$ RNAs were re-selected on 0.2 g of oligo(dT)-cellulose by adding LiCl (10 M) to the eluate for a final concentration of 0.5 M, and adding 40 mL of binding buffer. Samples were incubated as described above. Twice selected poly(A)$^+$ RNAs were eluted, extracted with phenol:choroform:isoamyl alcohol [phenol saturated with 100 mM Tris, pH 8.0 with 0.2% βME, 0.1% hydroxyquinoline and an equal volume of choroform:isoamyl alcohol 24:1 added] twice and chloroform:isoamyl alcohol alone once, and precipitated by adding 1/10 volume 3 M sodium acetate, pH 5.2 and 2.5 volumes of 100% ethanol.

Labeled RNA for use in crosslinking/label transfer were generated by *in vitro* runoff transcription using either T7 RNA polymerase (BRL) or SP6 RNA polymerase (BRL). Linearized DNA templates (1μg) were incubated at 37°C for 90 minutes in the presence of transcription buffer [(40 mM Tris-HCl, pH 8.0, 25 mM NaCl, 8 mM $MgCl_2$, 2 mM spermidine for T7) or (40 mM Tris-HCl, pH 7.9, 6 mM $MgCl_2$, 2 mM spermidine for SP6) , 10 mM DTT, 1U/μL RNasin (Promega), 0.5 mM ATP, 0.5 mM CTP, 0.02 mM GTP, 0.02 mM UTP, and 0.5 mM m_7GpppGTP caps (Pharmacia)], 40 μCi [α^{32}P]-UTP or -GTP (800Ci/mmol) and 1 μL of enzyme in a 50 μL reaction mix. To this, 15 μg yeast tRNA, 1 μL RQ DNase (Promega), and 1U/μL RNasin was added and incubation was continued an additional 10 minutes at 37°C. Products were extracted with phenol:chloroform:isoamyl alcohol and the organic phase was re-extracted with an equal volume of TE + 0.1% SDS. Both aqueous phases were combined and extracted with chloroform:isoamyl alcohol followed by two precipitations with an equal volume of 4M ammonium acetate and 2.5 volumes of 100% ethanol. After resuspending labeled RNA in 5 μL of DEPC- treated water, 10 μL formamide loading dye was added and samples were heated to 65° C for 10 minutes. Samples were cooled on ice for 2 minutes and then loaded onto an 8 – 10 % polyacrylamide/6M urea gel buffered with 1X TBE buffer (90 mM Tris-borate 2.5 mm EDTA). The gel was run for approximately 1500 volt-hours and labeled RNAs were excised, crushed and eluted for 2 hours in elution buffer (500 mM ammonium acetate, 10 magnesium acetate, 0.1 mM EDTA, 0.1% SDS). After running through a spin column [Ultra-Free MC 0.45μm filter (Millipore, Bedford MA)] to remove the polyacrylamide, the samples were extracted once with phenol:chloroform:isoamyl alcohol, once with chloroform:isoamyl alone, and

precipitated with an equal volume of 4M ammonium acetate and 2.5 volumes of 100% ethanol. Samples were stored at -80° C as an ethanol precipitate until ready for use.

Double-stranded DNA probes were labeled using a commercial random-primed labeling Kit (BRL) according to the manufacturer's protocol. To end label DNA oligonucleotides $(CAG)_{10}$, 200 ng of DNA in 16.5 uL of water was denatured at 70°C for 1 minute followed by quick cooling on ice. To the denatured DNA, 2.5 μL of 10X kinase buffer (0.5 M Tris-HCl, pH 7.6, 0.1 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine, 1 mM EDTA), 5 μL [γ^{32}P]- ATP (6000 Ci/mmol) and 1 μL of T4 polynucleotide kinase (BRL) were added and incubated at 37°C for 30 minutes. The reaction was terminated by adding EDTA to 15 mM followed by purification over Sephadex G25.

RNA Blot Analysis

Polyadenylated RNAs were purified from normal (HH and 3629) and DM lymphoblast cell lines (3986, 3756, 3696) as described above. Poly(A)$^+$ RNAs were denatured with glyoxal/DMSO, fractionated on a 1.0 % agarose gel (Sambrook et al., 1989) and transferred onto Hybond N$^+$ (Amersham) in 20X SSC. Blots were hybridized with a [α^{32}P]-dCTP random primed labeled (BRL) Bam HI/Hind III fragment of the DMPK gene (MTPK.2) or with a γ[^{32}P]ATP end labeled oligonucleotide $(CTG)_{10}$ at 65° C in 0.25 M $NaPO_4$, pH 7.4, 7 % SDS, 1 % BSA. Blots were washed twice in 2X SSC, 0.1% SDS at room temperature for 15 minutes and then in 0.5X SSC, 0.1 % SDS at 65° C for 30 minutes and were visualized by autoradiography.

Plasmid Constructs

The MTPK.2 plasmid was constructed by subcloning a BamH I-Hind III fragment (nt 2212-2849, DDBJ/EMBL/GenBank accession no. M87312) into pSP72 (Promega). Mutant DMPK plasmids (MTPK.8-4 containing 6 CTG repeats, MTPK.8-16 containing 54 CTG repeats, and MTPK.8-6 containing 90 CTG repeats) were created by PCR mutagenesis using flanking oligonucleotides, $(CTG)_{10}$ and $(CAG)_{10}$ oligonucleotides. Briefly, three separate PCR reactions were performed. Reaction A contained a DMPK specific 5' primer (MSS94) and an SP6 primer and MTPK.2 as a template. The 3' primer contains a single mutation that creates an EcoR I site just upstream of the (CTG) repeat. Reaction B contained DMPK specific 3' (MSS95) primer and a T7 primer and MTPK.2 as a template. The 5' primer contains a single mutation that creates a BamH I site just downstream of the (CTG) repeat. Reaction C contained only $(CAG)_{10}$ and $(CTG)_{10}$ oligonucleotides. All reactions were performed in a Perkin Elmer 9600 series Thermocycler using 3 temperature PCR (94° C for 30 seconds, 52° C for 30 seconds, and 72° C for 45 seconds) for 25 cycles in the presence of 1X PCR buffer (20 mM Tris-HCl, pH 8.4, 50 mM KCl, 1.5 mM $MgCl_2$, 0.2 mM dNTPs) 1 pmol/μL of each primer and 10 ng of template DNA. After completion of the initial PCR reaction, a second reaction was performed using products from either reaction A or B combined with products from reaction C. Finally, the two products from the A + C or B + C reactions were combined and amplified for an additional 25 cycles. Products were digested with Sma I and Hind III and cloned into pSP72 to create the MTPK.8 series of plasmids.

The DMPK clone deleted of CTG repeats (MTPK.10) was created by PCR mutagenesis similar to that described above using overlapping oligonucleotides (MSS63

and MSS64) that lacked CTG repeats. All clones were fully sequenced prior to use as templates for transcription. All MTPK constructs are cloned into pSP72 (Promega). Full length RPL14 clones in Bluescript were recloned into pSP72 at EcoR I and Xho I. pCTG10.2 was prepared using complementary oligonucleotides with 10 (CTG) or (CAG) repeats flanked by a EcoR I and BamH I site. The pCTG54 and pCTG90 were prepared by digesting MTPK.8-16 and MTPK.8-6 with EcoR I and BamH I and subcloning the resulting fragment into pSP72. The pCTG10.2, pCTG54, and pCTG90 contain identical flanking sequence and were verified by sequencing. The pThCTG11, pThCTG20, pThCTG35, pThCTG71 and pThCTG97 plasmids were kindly provided by Charles Thornton at the University of Rochester.

The pTAR and pmTAR clones were prepared by annealing primers MSS611 and MSS612 or MSS613 and MSS614 respectively followed by digestion with Xho I and Hind III. The resulting fragments were cloned into pSP72 and verified by sequencing. The TAR and mTAR (BL234) are identical to those described in Gatignol et al. (1991). For in vitro transcription, the clones were linearized with Hind III and the reactions carried out in the presence of SP6 RNA polymerase (BRL).

Photocrosslinking and Immunopurification

For RNA-protein photocrosslinking *in vivo*, HeLa S3 cells were grown in DMEM supplemented with 10% calf serum and 1% P/S to subconfluent densities. Cells were washed with ice-cold PBS and irradiated with UV light (Stratalinker, Stratagene) for 5 minutes in 5 mL of PBS at 4° C. Polyadenylated RNPs were isolated by lysing cells in 2.5 mL of lysis buffer (20 mM Tris-Cl [pH7.4], 1 mM EDTA, 50 mM LiCl, 1% sodium

dodecyl sulfate [SDS], 1% 2-mercaptoethanol, 1mg/mL of heparin, 10 mM vanadyl-adenosine, 1ug/mL leupeptin and pepstatin) per 10 cm plate and UV-crosslinked poly(A)$^+$ RNPs were isolated by oligo(dT) cellulose chromatography. Poly(A)$^+$ RNA was digested using RNase A and RNase T1 and proteins were analyzed by fractionation on SDS-PAGE and immunoblot analysis as described above.

For *in vitro* RNA binding studies, plasmids containing the 3'-UTR regions of the DMPK (MTPK.2, MTPK.8-4, MTPK.8-6, MTPK.8-16, and MTPK.10 linearized with Hind III), actin [pSP6γ-actin, linearized with BamH I], and RPL14 (RPL14.1 and RPL14.2 linearized with Xho I) genes were transcribed *in vitro* in the presence of [^{32}P]UTP or [^{32}P]GTP and RNAs were purified by denaturing gel electrophoresis. Following incubation of the labeled RNAs (20-50 fmoles) in a 25 μL reaction mix (11uL HeLa cell nuclear extract, 20 mM HEPES, pH 7.6, 1.3 mM $MgCl_2$, 1.5 mM ATP, 20 mM creatine phosphate) at 30° C for 10 min, 5 ug tRNA were added, samples were exposed to UV light (Stratalinker) for 5 min, and RNAs were digested with 2.5 μg RNase A (30 min at 37°C). Both total and immunopurified proteins photocrosslinked to RNAs were detected by label transfer/autoradiography following SDS-PAGE. Total protein samples fractionated by SDS-PAGE corresponded to 7.5 μl of the initial 25 μl reaction. Since the hnRNP C proteins crosslink more efficiently than other hnRNPs, the amount of the crosslinked reaction volume used for immunopurification varied from 25 μl (for mAb 4F4) to 190 μl (mAb 3B1). Immunopurification was performed at 4° C for 20 min essentially as described previously (Datar et al, 1993) except that Protein G-Sepharose was used and crosslinked samples were treated at 100° C in 1% SDS prior to dilution in

PBS containing 1mM EDTA, 1% Triton X-100, 0.5% deoxycholic acid, 0.1% SDS, 0.5% aprotinin.

Large-scale crosslinking/purification of EXP proteins for the production of anti-EXP antibodies was carried out using plasmid pCTG90A which is identical to pCTG90 except for a 21 nucleotide (A) stretch that was cloned at Hind III and Xho I using primers MSS127 and MSS128. Transcription was carried out by using 25 μg of linearized (Xho I) plasmid DNA in transcription buffer (80 mM HEPES-KOH, pH 7.5, 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT) with 3 mM rNTPs, 14.4 units of T7 RNA polymerase (BRL), 1 unit/μL rRNAsin (Promega), and 0.4 units/μL pyrophosphatase (Amersham) added in a 500 μL reaction. The reaction was incubated for 2.5 hours at 37° C and the RNA was purified over a Sepharose G-25 column. The RNA was extracted twice with phenol:chloroform:isoamyl alcohol and twice with chloroform:isoamyl alcohol followed by salt precipitation. This method produced approximately 2 mg of RNA per mL of reaction. RNA was resuspended in DEPC treated water and 100 μg of RNA was used for each crosslinking reaction. Large scale crosslinking was carried out as described above except that 100 μg of CUG90A RNA was used with 11mL HeLa cell nuclear extract (20 mM HEPES, pH 7.6, 1.3 mM $MgCl_2$, 1.5 mM ATP, 20 mM creatine phosphate) in a 25 mL reaction volume. Samples were incubated and crosslinked as described above. Crosslinked RNA/RNPs were isolated on an oligo(dT) cellulose column as described for in vivo crosslinking above and precipitated with 0.2 M LiCl and 2.5 volumes of 100% ethanol. RNA/RNPs were resuspended in 100 μL of RSB-100 and injected into BALB/c mice for the production of antibodies.

RESULTS

The main focus of this project was to understand the roles that hnRNPs play in pre-mRNA processing and how this relates to human disease. The initial isolation of hNab50 and its characterization are described followed by its identification as the $(CUG)_8$ RNA-binding protein. Data supporting an RNA-dominant mutation model and the possible involvement of hNab50 in myotonic dystrophy are presented. Finally the identification of the EXP proteins is described and the implications to DM disease are discussed.

Proteins that Interact with Nab2p are Involved in mRNA Export and Polyadenylation

Immediately following transcription by RNA polymerase II, nascent pre-mRNA transcripts become associated with the heterogeneous nuclear ribonucleoproteins (hnRNPs) and small nuclear ribonuclear proteins (snRNPs). These transcripts then undergo a multitude of processing events including capping, splicing, and polyadenylation to convert pre-mRNAs to mRNAs which are subsequently exported from the nucleus and translated into proteins in the cytoplasm. Although metazoan hnRNPs have been under intense study over the last twenty years, their role in the maturation of pre-mRNAs is not well understood. To better understand the role that hnRNPs play in the biogenesis of mRNAs, the characterization of these proteins was undertaken in the yeast, *Saccharomyces cerevisiae* (Anderson et al., 1993; Wilson et al., 1994). Studies in

S. cerevisiae have allowed the biochemical and genetic evaluation of the functions of these proteins.

Nab2p is one of four nuclear polyadenylated RNA-binding proteins that were originally identified in a screen to isolate hnRNPs from *Saccharomyces cerevisiae* using an in vivo UV crosslinking strategy (Anderson et al., 1993; Wilson et al., 1994). Nab2p crosslinks to poly(A)$^+$ RNA in vivo and is primarily nuclear in its subcellular distribution, suggesting that its primary function is in the nucleus. Nab2p is essential for viability and is required for correct pre-mRNA processing at the level of polyadenylation and nuclear mRNA export. Both *nab2Δ* deletion mutants and temperature-sensitive mutants exhibit an increase in poly(A) tail length and accumulate poly(A)$^+$ RNA in the nucleus (Anderson et al., 1993; Anderson 1994). The *nab2* mutant phenotype suggested that Nab2p was an important component of both polyadenylation and mRNA export and that these two processes were tightly linked in vivo. In addition, recent work has shown an interaction between Nab2p and yTRN1/Kap104p, the yeast homolog of the nuclear transport protein transportin (Aitchison et al., 1996; Truant et al., 1998; Siomi et al., 1998).

Although Nab2p is one of the major hnRNPs found in yeast, it does not have a direct structural homolog in metazoans. Since hnRNPs often form homo-oligomers in vitro, Nab2p interactive proteins in human cells were sought using the two-hybrid system. A cross-species two-hybrid screen was performed to determine the parallel pathways that Nab2p might be involved in and to isolate a Nab2p homolog in humans. A human HeLa cDNA library was screened using the full length Nab2 protein as bait. Of ~750,000 transformants, 18 grew in the absence of histidine (His+) and 13 were positive

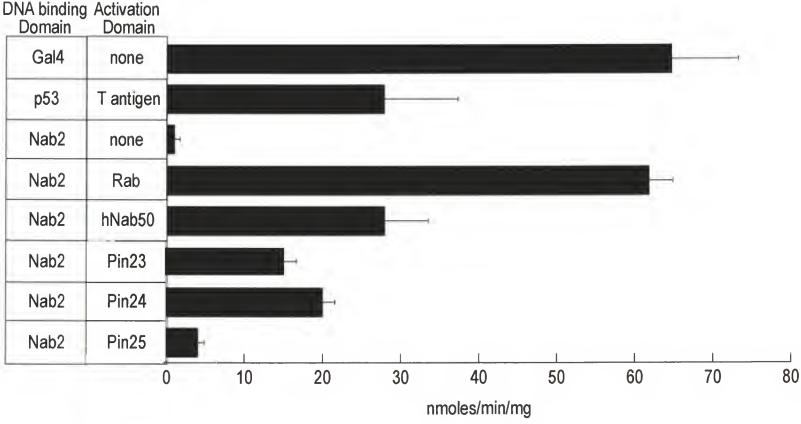
Table 4: Proteins that interact with Nab2 (Pins)

Pin#	Identity	activity	# of isolates	reference
Pin21	Rab/Rip	+++	3	Bogerd 1995
Pin22	hNab50	++	1	Timchenko 1996
Pin23	SAF-B	++	1	Renz 1996
Pin24	hnRNP D	++	4[a]	Kajita 1995
Pin25	hMCM2	+	1	Todorov 1994
Pin26	Pab II	++	3[a]	Wahle 1991
Pin27	Siah BP	+	1	Hu 1996
Pin28	SC35	+	1	Fu 1992
Pin29	HSC10-II	+	1	Nothwang 1994

HF7c *MATa, ura3-52, his3-Δ200, lys2-801, ade2-101, trp1-Δ901, leu2-3, -112, gal4-Δ542, gal80-Δ538, LYS2::GAL1-HIS3, URA3::(GAL4 17-mers)$_3$-CYC1-lacZ*

a. two of the PAB II and one of the hnRNPD clones were isolated as human cDNA contaminants of a yeast two-hybrid
library that was screened at a later date.

Figure 3. Isolation of hNab50. The β-galactosidase activities of the His+/β-gal+ yeast strains isolated using a two-hybrid interaction system (Clontech) are shown. Positive controls include intact Gal4p and the interaction between human p53 and SV40 large-T antigen.


DNA binding Domain
Activation Domain
Gal4
none
p53
T antigen
Nab2
none
Nab2
Rab
Nab2
hNab50
Nab2
Pin23
Nab2
Pin24
Nab2
Pin25
0
10
20
30
40
50
60
70
80
nmoles/min/mg

for β-galactosidase activity. Proteins that interact with Nab2 (PINs) are listed in Table 4 and the results of a quantitative β-galactosidase liquid assay for several of the clones are depicted in Figure 3.

The Pin21 clone had the highest β-galactosidase activity and was isolated three times in the screen. A protein homology (BLAST; Altschul, et al., 1990) search revealed Pin21 to be identical to human Rab/hRip (Rev/Rex activation domain binding protein or human REV interacting protein). The Rab/hRip protein is a cellular protein that was originally isolated as interacting with a leucine-rich "activation domain" of HIV Rev and HTLV-1 Rex proteins (Stutz et al., 1995; Bogerd et al., 1995) and is involved in nuclear export of proteins. Interaction of Rab/hRip with Nab2, an mRNA binding protein, suggested that Nab2 might facilitate the export of mRNAs from the nucleus.

The Pin26 clone was isolated three times and was found to be identical to the polyadenylation factor, PAB II [poly (A) binding protein II]. PAB II is a processivity factor for poly (A) polymerase and is involved in the regulation of poly(A) tail length in conjunction with CPSF (Wahle 1991, 1995). Yeast also restrict poly(A) tail length *in vivo*, however, no structural PAB II homologue exists in yeast. The fact that *nab2* mutants display a long tail phenotype and that Nab2p interacts with human PAB II in the two-hybrid system supports the hypothesis that Nab2 is directly involved in polyadenylation, possibly at the level of tail length regulation.

Another RNA-binding protein, hnRNP D, was isolated four times and is designated Pin24 (Table 4 and Figure 3). Several forms of the hnRNP D proteins have been previously isolated and are components of the major hnRNP complex in HeLa cells (Kajita et al., 1995; Dreyfuss et al., 1993). Although primarily nuclear in its subcellular

localization, hnRNP D (AUF1) has been found to bind to AU-rich elements suggesting a role in cytoplasmic mRNA stability (Zhang et al., 1993). The first step of deadenylation-dependent decay of mRNAs is 3' →5' digestion of the poly(A) tail. Proteins that bind to AU-rich elements are thought to be important in regulating this rate-limiting step of mRNA decay (Decker and Parker, 1994). Interestingly, hnRNP D was also found to be a component of the α-stability complex found on the α-globin mRNA (Kiledjian et al., 1997). During red blood cell maturation, all but a subset of mRNAs are degraded thus making room for the high level expression of the globin gene products. Cis-elements within the 3'-UTR of α-globin have been found to prevent its degradation in what appears to be a "default" pathway. The α-stability complex is a complex of proteins that bind to this element, suggesting a function in stabilizing this mRNA (Russell et al., 1997).

The scaffold attachment factor B (SAF-B), which was isolated once in the screen, was originally identified as one of several proteins that binds S/MAR DNA elements. Many investigators believe that the topological organization of chromatin facilitates gene expression through scaffold attachment and matrix attachment regions (Bode, et al., 1995). Recently, SAF-B has been found to interact with RNA polymerase II and several SR proteins, as well as colocalizing with SC35 by indirect immunofluorescence (Nayler, et al., 1998). Interestingly, the Pin28 protein was isolated once and is identical to the splicing factor SC35 (Fu and Maniatis, 1992), which belongs to the class of SR proteins containing both a CS-RBD and an RS domain. The SC35 protein is believed to play a role in both constitutive and alternative pre-mRNA splicing (Fu and Maniatis, 1992; Fu, 1995). Pin25 was identified as hMCM2 (BM28), a human homolog to the yeast MCM2

protein (Todorov, et al., 1994). The MCM (mini-chromosome maintenance proteins) are factors involved in initiation of DNA replication and are required for cell cycle progression in yeast (Campbell, 1993). The Pin27 clone is identical to a protein found to interact with siah ring finger proteins, Siah BP (Hu, 1996). The siah family of proteins are mammalian homologs of the *Drosophila seven in absentia* (*sina*) gene. These proteins are thought to be involved in the regulation of cell fate by signal transduction pathways (Della et al., 1993). The siah binding protein has not been fully characterized, but has homology to several RNA binding proteins by protein alignment analysis and contains three putative RNA-binding domains (J. Miller and M.S. Swanson, unpublished observation).

Pin22 was isolated once and was found to be a novel protein, which contains three consensus RNA-binding domains (RBDs). The overall structure of this protein resembles a family of proteins that are involved in a variety of aspects of mRNA maturation and metabolism. Characterization of this protein, which has been designated hNab50, is described in the following section.

Characterization of hNab50: A Novel Human hnRNP

The hNab50 protein was originally isolated in the two-hybrid screen described above using the yeast hnRNP, Nab2p, as bait. This novel nuclear polyadenylated RNA-binding (Nab) protein contains three consensus sequence RNA-binding domains (CS-RBDs) and is structurally related to a family of proteins termed the *elav*-like RNA binding proteins or ELR proteins (Figure 4). The original *Drosophila* Elav (embryonic lethal abnormal visual system) protein is essential for viability and is involved in

Figure 4. hNab50 belongs to the Elav-like family of proteins. At the top, the common primary structure of the Elav-like family of proteins is depicted. Each contains two amino-terminal consensus sequence RNA-binding domains (CS-RBD) and one carboxy-terminal CS-RBD joined by a hinge region that varies between the different proteins. Many of the Elav-like proteins have been found to be neuronal specific, while others are ubiquitously expressed. Functions for these proteins range from regulated splicing to mRNA 3' end formation and poly(A) tail length control.

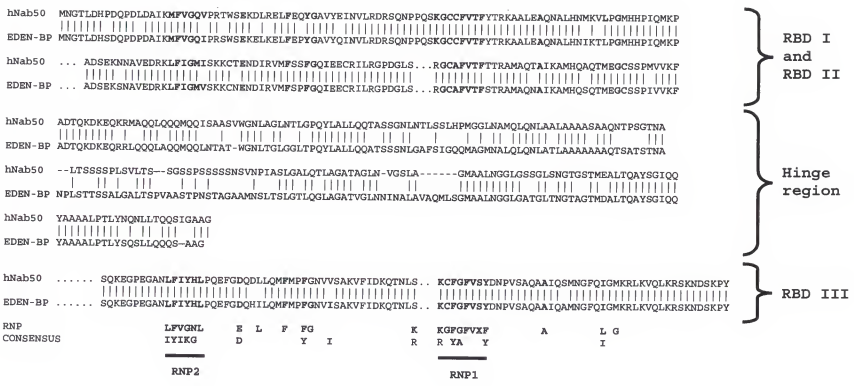
elav-like gene family

Organism/Protein	Expression Pattern	Proposed Function
Drosophila elav	neuronal	neuronal differentiation and mainteneance
Human HuD, HuC, Hel-N1	neuronal	neuronal differentiation; neuroendocrine tumors; bind to AREs in mRNA 3'-UTR; mRNA turnover
Human HuR	ubiquitous	bind to AREs in mRNA 3'-UTR; mRNA turnover
Human hNab50	ubiquitous	bind to (CUG)n repeats and DMPK mRNA
Xenopus EDEN-BP	oocyte/embryo	poly(A) tail length regulation

development of the *Drosophila* nervous system (Campos et al., 1985, Robinow and White 1988) and is also implicated in neuronal specific alternative splicing (Koushika et al., 1996). Protein sequence alignment analysis has revealed three proteins with high homology to hNab50. The Hetr3 protein (79% identity) is a human protein that was identified in a sequencing project of a fetal heart library and has not been further characterized (Hwang et al 1994). Although Hetr3 is remarkably similar to hNab50 at the amino acid level, its sequence at the nucleotide level is degenerate, particularly at the wobble base, indicating that it is derived from a different gene. A partial clone of a mouse protein, "mouse brain protein," is highly homologous to hNab50 with 98% amino acid identity (from residues 127-362) and was identified in a subtractive library screening project of the mouse brain (Kato, 1992). Mouse brain protein was found to localize to the neocortex and putamen by *in situ* hybridization but has not been further characterized. The protein with highest homology overall to hNab50, and the most interesting in terms of function, is EDEN-BP (embryonic deadenylation element binding protein; Paillard et al., 1997) with 88.4% amino acid identity (see Figure 5). The *Xenopus* protein binds to a GU-rich deadenylation element found in the 3'UTR of several maternal mRNAs. The EDEN element is required for post-fertilization deadenylation of a subclass of maternal RNAs. Elimination of EDEN abolishes both EDEN-BP binding and deadenylation. In addition, immunodepletion of egg extracts of EDEN-BP also abolishes deadenylation of these transcripts (Paillard et al., 1998).

Thus, hNab50 resembles proteins involved in several aspects of pre-mRNA/mRNA metabolism, particularly proteins that bind to the 3'UTR of mRNAs. It also is structurally related to hnRNP proteins, which are believed to be fundamental in the

Figure 5. Sequence comparison between hNab50 and EDEN-BP. The amino acid sequence of hNab50 is compared to the *Xenopus* homolog, EDEN-BP, which is 88% identical. The ribonucleoprotein consensus sequence RNA-binding domains (RBDs) are shown in brackets, as is the hinge region. The most conserved motifs within the RBDs, RNP1 and RNP2, are depicted at the bottom by underlining (Bandzuilis et al., 1989)

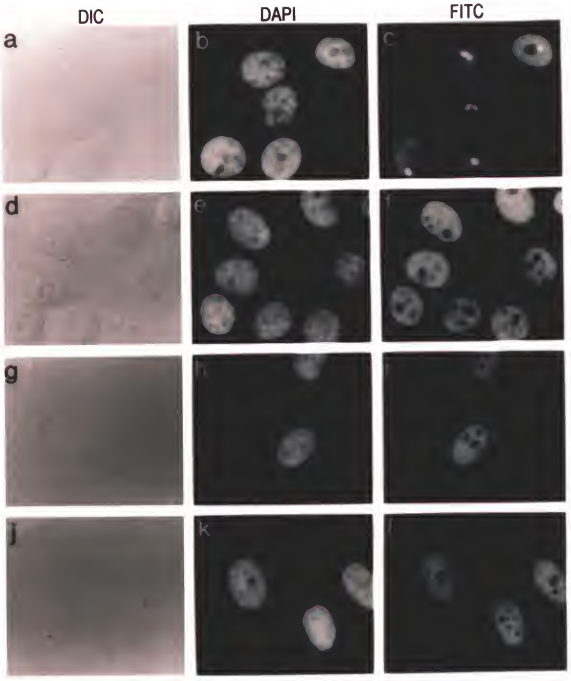

hNab50 MNGTLDHPDQPDLDAIKMFVGQVPRTWSEKDLRELFEQYGAVYEINVLRDRSQNPPQSKGCCFVTFYTRKAALEAQNALHNMKVLPGMHHPIQMKP
EDEN-BP MNGTLDHSDQPDPDAIKMFVGQIPRSWSEKELKELFEPYGAVYQINVLRDRSQNPPQSKGCCFVTFYTRKAALEAQNALHNIKTLPGMHHPIQMKP
hNab50 ... ADSEKNNAVEDRKLFIGMISKKCTENDIRVMFSSFGQIEECRILRGPDGLS...RGCAFVTFTTRAMAQTAIKAMHQAQTMEGCSSPMVVKF
EDEN-BP ... ADSEKSNAVEDRKLFIGMVSKKCNENDIRVMFSPFGQIEECRILRGPDGLS...RGCAFVTFSTRAMAQNAIKAMHQSQTMEGCSSPIVVKF
RBD I and RBD II
hNab50 ADTQKDKEQKRMAQQLQQQMQQISAASVWGNLAGLNTLGPQYLALLQQTASSGNLNTLSSLHPMGGLNAMQLQNLAALAAAASAAQNTPSGTNA
EDEN-BP ADTQKDKEQRRLQQQLAQQMQQLNTAT-WGNLTGLGGLTPQYLALLQQATSSSNLGAFSIGQQMAGMNALQLQNLATLAAAAAAAQTSATSTNA
hNab50 --LTSSSSPLSVLTS--SGSSPSSSSSSNSVNPIASLGALQTLAGATAGLN-VGSLA------GMAALNGGLGSSGLSNGTGSTMEALTQAYSGIQQ
EDEN-BP NPLSTTSSALGALTSPVAASTPNSTAGAAMNSLTSLGTLQGLAGATVGLNNINALAVAQMLSGMAALNGGLGATGLTNGTAGTMDALTQAYSGIQQ
Hinge region
hNab50 YAAAALPTLYNQNLLTQQSIGAAG
EDEN-BP YAAAALPTLYSQSLLQQQS--AAG
hNab50 ..... SQKEGPEGANLFIYHLPQEFGDQDLLQMFMPFGNVVSAKVFIDKQTNLS.. KCFGFVSYDNPVSAQAAIQSMNGFQIGMKRLKVQLKRSKNDSKPY
EDEN-BP ..... SQKEGPEGANLFIYHLPQEFGDQHILQMFMPFGNVISAKVFIDKQTNLS.. KCFGFVSYDNPVSAQAAIQAMNGFQIGMKRLKVQLKRSKNDSKPY
RBD III
RNP LFVGNL E L F FG K KGFGFVXF A L G
CONSENSUS IYIKG D Y I R R YA Y I
RNP2
RNP1

structuring of pre-mRNAs such that they are the correct substrates for subsequent processing and export from the nucleus. From these data, we hypothesized that hNab50 is an hnRNP, but that it differs from the major cellular hnRNPs by directing processing events for a subset of pre-mRNAs.

Subcellular Localization of hNab50

For hNab50 to be classified as a heterogenous nuclear ribonucleoprotein (hnRNP) it must be primarily nuclear in its subcellular distribution and must associate with poly $(A)^+$ RNA *in vivo*. To investigate the subcellular localization of hNab50, indirect immunofluorescence microscopy was performed on HeLa cells and human myoblasts using mAb 3B1 against hNab50 (Figure 6 c, i, and l). The hnRNP M proteins, which show a strong nuclear signal, were also localized as a control (Figure 6 f). The hNab50 protein was primarily nuclear in HeLa cells, normal and DM myoblasts (Figure 6 c, i, and l) as well as several other cell types tested (IMR-90, Hep2, A549 and lymphoblast; not shown). Interestingly, hNab50 accumulated in a peri-nucleolar region in HeLa cells and Hep2 cells (not shown) but these foci were absent in myoblasts (Figure 6 i and l), lymphoblasts, A549, and IMR-90 cells (not shown). The localization and quality of the peri-nucleolar foci in HeLa cells resembled that of hnRNP I and several Y RNAs which have been found to accumulate in a nuclear subcompartment structure known as the peri-nucleolar compartment or PNC (Ghetti et al., 1992; Matera et al., 1995; Huang et al., 1997). The PNC represents an area of accumulation of certain RNA polymerase III transcripts but it is devoid of Ro proteins, which normally bind to these transcripts. PNCs are more often found in highly transformed cell lines but can also be detected at

Figure 6. hNab50 is a nuclear RNA-binding protein. The hNab50 and hnRNP M proteins were localized within cells by indirect cellular immunofluorescence using either the 3B1 (c, i, l) or 1D8 (f) MAbs, respectively. HeLa (a-f), normal myoblasts (g-i) and DM myoblasts (j-l) are shown. The positions of the cells are shown by differential interference contrast (DIC) microscopy (a,d,g,j), and the chromosomal DNA by DAPI staining (b,e,h,k).


DIC
DAPI
FITC
a
b
c
d
e
f
g
h
i
j
k
l

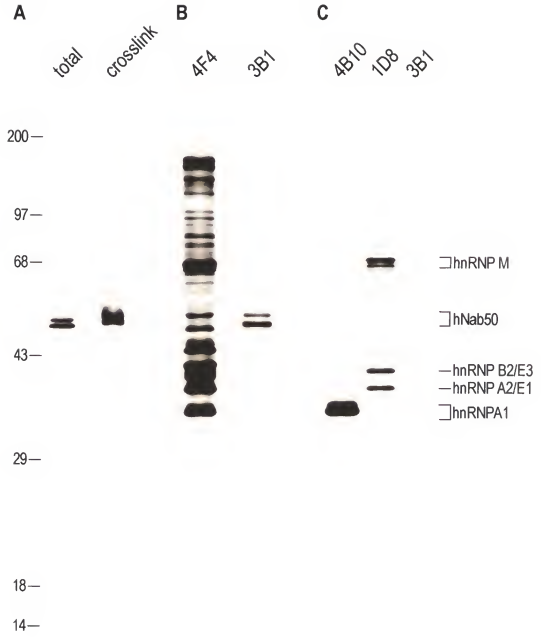
lower levels in immortalized lines (Matera et al., 1995; Huang et al., 1997). The function of the PNC is unknown, but may represent some stage of RNA biogenesis in RNA pol III metabolism. The fact that hnRNP I and, possibly, hNab50 were also found in this nuclear subcompartment may represent a binding specificity of these proteins, or a distinct function that has not yet been described.

hNab50 is a Poly(A)$^+$ RNA-Binding Protein

In vivo UV crosslinking has been extremely useful in identifying RNA-binding proteins that are in direct contact with RNA molecules in the intact cell (Mayrand et al, 1981; Dreyfuss et al., 1984). Ultra-violet light photo-activates the RNA and allows it to react with proteins in close proximity to form covalent bonds that are resistant to detergent and high salt (Greenberg, 1980). To determine if hNab50 associated with poly(A)$^+$ RNA *in vivo*, HeLa cells were UV irradiated and then lysed in the presence of protease inhibitors and poly(A)$^+$ RNA/RNPs isolated. Subsequent immunoblot analysis of the crosslinked proteins revealed that hNab50 was associated with poly(A)$^+$ RNA *in vivo* (Figure 7A). Thus, since hNab50 was primarily nuclear in its subcellular localization and it associated directly with poly (A)$^+$ RNA in vivo, it was classified as an hnRNP.

Since hNab50 was an authentic hnRNP, we wanted to test whether it copurified with the major hnRNP complex in HeLa cells (Figure 7B). The hnRNP complex consists of >20 proteins that can be copurified using MAbs directed against different proteins found in the complex (Dreyfuss et al., 1993). The proteins of the major hnRNP complex are among the most abundant in actively growing cells. This is the major complex of

Figure 7. The hNab50 protein is associated with poly(A)$^+$ RNA in vivo but fails to co-immunopurify with hnRNP complexes. **(A)** hNab50 is associated with poly (A)$^+$ RNA in vivo. Total HeLa cell proteins (total) or proteins photocrosslinked to poly (A)$^+$ RNAs in vivo (crosslink) were immunoblotted with MAb 3B1. The decrease in relative mobility of hNab50 in the crosslink lane is due to crosslinked nucleotides which remain following nuclease digestion. Sizes are indicated in kilodaltons. **(B)** A monoclonal antibody against hNab50 fails to immunopurify the hnRNP complex. HeLa cells were labeled with [^{35}S]methionine and hnRNP complexes immunopurified using MAb 4F4 against the hnRNP C proteins (4F4). Parallel immunopurifications were performed using MAb 3B1 against hNab50 (3B1). **(C)** hNab50 is not a major component of the immunopurified hnRNP complex. HnRNP complexes were isolated as described in (B) using MAb 4F4. RNA-protein complexes were then dissociated by 1% SDS and boiling followed by dilution to 0.1% SDS. Monoclonal antibodies were then used to immunopurify hnRNP A1 (4B10), hnRNP M (1D8) and hNab50 (3B1). Under these conditions, MAB 1D8 also immunopurifies proteins that co-migrate with the hnRNP A2/E1 and B2/E3 proteins.


A
total
crosslink
B
4F4
3B1
C
4B10
1D8
3B1
200
97
68
43
29
18
14
hnRNP M
hNab50
hnRNP B2/E3
hnRNP A2/E1
hnRNPA1

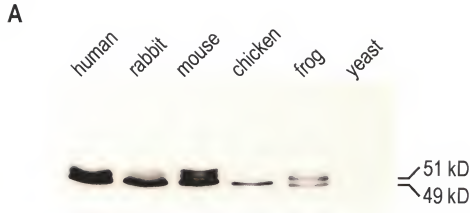
proteins found in association with nuclear pre-mRNA/mRNA and is believed to be responsible for structuring the RNA in such a way that it can act as a substrate for further processing events. To test if hNab50 was a component of the major hnRNP complex, HeLa cells were labeled with ^{35}S-methionine and nucleoplasm was isolated under mild conditions that preserve protein/protein and protein/RNA interactions. Complexes were then immunoprecipitated using MAb 4F4 against hnRNP C or MAb 3B1 against hNab50. The MAb 4F4 immunopurified the entire complex while 3B1 only immunopurified hNab50 (Figure 7B). Antibodies to some components of the hnRNP complex are unable to purify the entire complex, presumably because their epitopes are unavailable or their association is not stable enough to isolate the whole complex (Datar et al., 1993). To test this possibility, the hnRNP complex was first isolated using MAb 4F4 and then these isolated complexes were denatured and subjected to a second immunopurification using MAb 3B1, 4B10 against hnRNP A1 or 1D8 against the M proteins (Figure 7C). HnRNP A1 and M were both efficiently immunopurified under these conditions, but hNab50 was not. This demonstrated that hNab50 was not a stable component of the major hnRNP complex. It is possible that hnRNP complexes containing hNab50 are not soluble under the conditions employed or that it is loosely associated and is lost during the purification process. Since hNab50 associated with poly(A)$^+$ RNA in vivo, it may associate with a subset of poly(A) RNAs in a sequence-specific manner. Studies carried out in *Drosophila* and in amphibian oocyte lampbrush chromosomes have revealed a unique assemblage of hnRNPs on a various nascent pre-mRNA transcripts (Pinol-Roma et al., 1989; Matunis et al., 1993). These studies suggest that the binding of hnRNPs to nascent transcripts occurs with a stoichiometry that is representative of the sequences found in

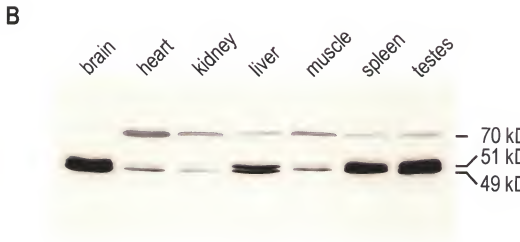
that transcript. In addition, several of the ELR proteins have been found to have sequence-specific binding sites on particular mRNAs (Jain et al., 1997; Chung et al., 1997; Joseph et al., 1998).

hNab50 is Immunologically Conserved

With such closely related homologs in both mouse and frog, we wanted to investigate a variety of different organisms for proteins immunologically related to hNab50 (Figure 8a). Total cellular proteins were prepared from cell lines derived from human (HeLa), rabbit (RK13), mouse (3T3), chicken (CEF), frog (XL1), and budding yeast (BJ926) and analyzed by immunoblotting using MAb antibody 3B1 against hNab50. All vertebrate species tested revealed proteins of similar molecular weight that were reactive with MAb 3B1. Yeast and *Drosophila* (not shown) did not contain immunoreactive proteins. Analysis of 3B1-reactive proteins in different mouse tissues revealed that these proteins were ubiquitously expressed and that several different forms existed (Figure 8b). The 49 kD form seen in human HeLa cells was seen in all tissues tested while both the 49 and 51 kD forms that were seen in brain, liver, spleen and testes. In addition, a novel 70 kD protein was detected in all tissues but brain. It was not clear whether this was an alternative form of hNab50 or a different protein that was immunoreactive with 3B1 antibody. These data strongly suggested that hNab50 has direct homologs in many different species and likely performs a highly conserved and essential task.

Figure 8. Proteins immunologically related to hNab50. (A) Immunoblot analysis using MAb 3B1 against hNab50. Total cellular proteins were isolated from either human (HeLa), rabbit (RK13), mouse (NIH3T3), chicken (CEF), frog (XL1) or budding yeast (BJ926) cells grown in culture. **(B)** Total cellular protein extracts derived from a variety of mouse tissues were immunoblotted using MAb 3B1 against hNab50. Proteins that migrate at the same molecular weight as the human protein can be seen in all tissues. In addition, a 70 kD form that is not detected in HeLa cells, can also be seen.


A
human
rabbit
mouse
chicken
frog
yeast
51 kD
49 kD

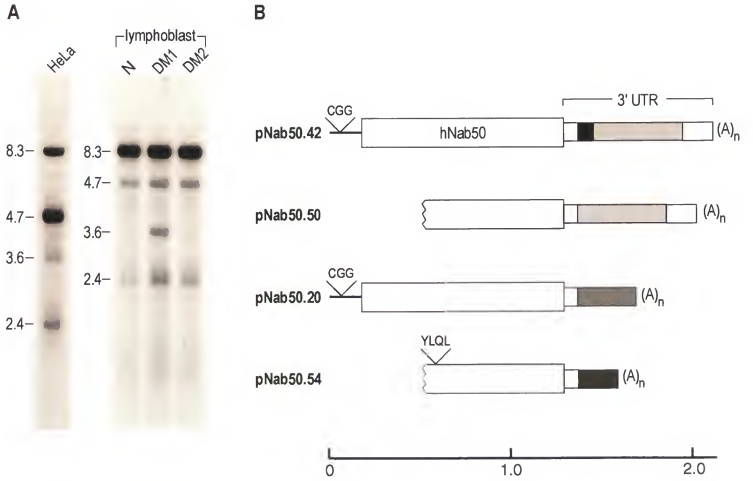

B
brain
heart
kidney
liver
muscle
spleen
testes
70 kD
51 kD
49 kD

Variable Sequences Found in hNab50 3'-UTR

Many different variants of hnRNP proteins have been found to result from either alternative splicing of the RNA transcript or from post-translational modifications of the protein product (Dreyfuss et al., 1993). The pattern seen on immunoblot for hNab50 suggested that it may also be modified in such a manner. To further investigate possible alternative splicing of hNab50 transcripts, RNA blot analysis of poly(A)$^+$ RNA from HeLa cells was performed using a restriction fragment derived from the original hNab50 two-hybrid clone. Four major RNA species were present with a 4.7 kb band as the most abundant in HeLa cells and the 8.3 kb band is predominant in both normal and DM patient lymphoblasts (Figure 9A). We knew from the immunoblot analysis that hNab50 was ubiquitously expressed, and that different forms of the protein were abundant in different tissues. The RNA blot analysis supported this data by showing the same RNA species in both HeLa and lymphoblast cells, but with different distributions.

To try and resolve the question of alternative splicing of hNab50, several different libraries were screened by hybridization using a restriction fragment derived from the original hNab50 two-hybrid clone. Full length clones were obtained from both an osteosarcoma cDNA library and a HeLa cDNA library. Full length clones were particularly difficult to isolate due to an extremely GC rich segment in the 5' UTR of the gene (Figure 9B). Several different potential alternative splice sites were determined by examining the isolated clones. Interestingly, nearly all of the differences seen between the clones occurred in the 3' UTR and not in the coding region. Only one small stretch of four amino acids in the coding region was found to be different between the HeLa and

Figure 9. Different hNab50 mRNAs vary in the 3'-UTR region. **(A)** RNA blot analysis of poly(A)$^+$ RNAs from HeLa, one normal (N) and two DM (DM1, DM2) lymphoblast cell lines. Sizes are indicated in kilobases. **(B)** Diagram of the structures of several HeLa cell hNab50 cDNA clones. The 5'-UTR is designated by a thick line and contains 14 (CGG) repeats. The hNab50 open reading frame (ORF) is illustrated by an open box with the position of the tetrapeptide insert highlighted (YLQL). The position of the 3'-UTR is indicated and contains a variable DNA sequence between cDNA clones (shown by open, black, or grey boxes). The only full-length clones isolated were 2.4 kb in size.


A
HeLa
lymphoblast
N
DM1
DM2
8.3
4.7
3.6
2.4
8.3
4.7
3.6
2.4
B
3' UTR
CGG
pNab50.42
hNab50
(A)n
pNab50.50
(A)n
CGG
pNab50.20
(A)n
YLQL
pNab50.54
(A)n
0
1.0
2.0

osteosarcoma clones (Figure 10B). This suggested that the 3'-UTR of hNab50 was important for post-transcriptional regulation of expression.

Thus, we concluded that hNab50 was an hnRNP which likely had transcript-specific binding properties. Its similarity to the *elav*-like proteins suggested that it might be a 3'-UTR binding protein, which regulated pre-mRNA processing. The interaction of hNab50 with yeast Nab2p and its striking similarity to EDEN-BP suggested that it might be involved at the level of polyadenylation or mRNA export. During this time, Dr. Lubov Timchenko and Dr. C. Thomas Caskey had purified a CUG-binding activity that was approximately 50 kD in size. Since the defect in the DMPK transcript was in the 3'-UTR and there was evidence that both polyadenylation and mRNA export of the DMPK RNA was affected in the disease state, I decided to contact these investigators and initiate a collaboration to test if their CUG-binding protein might be hNab50.

hNab50 is the CUG Repeat RNA-Binding Protein

The identification of a CUG repeat binding activity was initially carried out in the laboratory of our collaborators, Dr. Lubov Timchenko and Dr. C. Thomas Caskey (Timchenko et al., 1996a). HeLa cell subcellular fractions were tested for binding activity using an end-labeled RNA oligonucleotides consisting of eight (CUG) repeats and an electromobility shift assay (EMSA). In total cellular extracts, two activities were identified that shifted the $(CUG)_8$ RNA probe (Timchenko et al., 1996a). These activities, designated CUG-BP1 and CUG-BP2, were then analyzed further using column chromatography. A size selected fraction, p46, was further fractionated on a DEAE-Sepharose column using NaCl step gradient. Each protein fraction was analyzed for

$(CUG)_8$ binding activity by EMSA. The CUG-BP1 activity eluted between 0.2 and 0.3 M NaCl, while CUG-BP2 activity was found in the flow-through. The partially purified CUG-BP1 and CUG-BP2 activities were present in both the nucleus and the cytoplasm and were estimated to be between 40 and 50 kD in molecular weight. A collaborative effort was established to determine if hNab50, which I had previously isolated and characterized from HeLa cells, was responsible for the (CUG)-binding activity.

Purified CUG-BP1 and CUG-BP2 were tested by immunoblot using MAb against hNab50 and two abundant hnRNPs, which were within the same size range. For example, the hnRNP C proteins (Swanson et al., 1987) are two abundant hnRNPs with an apparent molecular weight of 41 and 43 kD by SDS-PAGE. The hnRNP C2 protein is identical to C1 except for an additional 13 amino acids located in the middle of the protein. The function of these proteins is not understood although they have been implicated in pre-mRNA splicing (Choi et al., 1986) and they are one of the primary components of the major hnRNP complex (Beyer et al., 1977). The hnRNP A/B proteins are also extremely abundant hnRNP proteins with molecular weights ranging between 34 and 40 kD (Dreyfuss et al., 1993). Three proteins, hnRNP A1, A2 and B1 are all structurally related and are likely alternatively spliced variants of the same gene (Burd et al., 1989). Functionally, the hnRNP A/B proteins are important in alternative pre-mRNA splicing and are known to shuttle between the nucleus and the cytoplasm (Harper and Manley, 1992; Mayeda and Krainer, 1992; Piñol-Roma and Dreyfuss, 1992).

Monoclonal antibodies directed against each of these proteins were used to detect proteins by immunoblot analysis. While mAb 4F4 against hnRNP C and 4B10 against hnRNP A1 did not recognize the purified fractions by immunoblot, MAb 3B1 directed

against hNab50 did give a positive signal against both CUG-BP1 and CUG-BP2 (Timchenko et al., 1996b). This suggested that the two activities were alternatively spliced or post-translationally modified forms of the same protein. In addition, MAb 3B1 supershifted and neutralized CUG-BP complexes bound to labeled $(CUG)_8$ RNA and recombinant hNab50 was able to bandshift the $(CUG)_8$ probe in the presence of competitor RNAs (Timchenko et al., 1996b). Therefore, hNab50 is the protein responsible for the CUG binding activity detected by EMSA analysis.

CUG-BP/hNab50: Disease-Associated Changes in DM

The isolation and characterization of a CUG-binding protein allowed us to evaluate the role of this protein in the pathogenesis of myotonic dystrophy. We first wanted to determine if there were any differences in CUG RNA-binding activity between DM and normal patient cell extracts. EMSA analysis was performed using an end-labeled $(CUG)_8$ probe and nuclear and cytoplasmic fractions derived from several DM and normal patient lymphoblasts and myoblasts (Timchenko et al., 1996b). Although both CUG-BP1 and CUG-BP2 activity was seen in the nuclear and cytoplasmic fractions in the normal and DM cell lines, their distribution was altered in the DM lines (Table 5). The majority of CUG-BP1 activity was seen in the cytoplasmic fraction of normal cells although significant activity was also is present in the nuclear fraction. In DM patient cells, however, there was a significant increase in CUG-BP2 activity, and a decrease of CUG-BP1 activity, in the nuclear fraction. It was later shown that two forms of hNab50, CUG-BP1 and CUG-BP2, result from differential phosphorylation (Roberts et al., 1997).

Table 5 CUG-BP/hNab50: Disease-Associated Changes in DM[a]

	cytoplasmic fraction	nuclear fraction
Normal		
CUG-BP1	+++[b]	++
CUG-BP2	+	+++
DM		
CUG-BP1	+++	+
CUG-BP2	+	++++

Summary of changes in CUG-BP1 and CUG-BP2 activity in DM and normal lymphoblast cell lines. Cells were divided into cytoplasmic and nuclear fractions as described in materials and methods. EMSA analysis was performed using a $(CUG)_8$ end-labeled RNA probe.

[a] Data derived from Timchenko et al., (1996b).

[b] Plus (+) symbols refer to the bandshifting activity seen for each fraction.

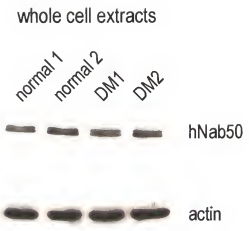


Figure 10. No alteration in total protein concentration of hNab50 between DM and normal cells. Total cellular proteins were prepared from DM and normal lymphoblast cells. Proteins were separated on an SDS-PAGE gel and proteins were detected by immunoblot using either MAb 3B1 against the hNab50 protein or C4 against actin.

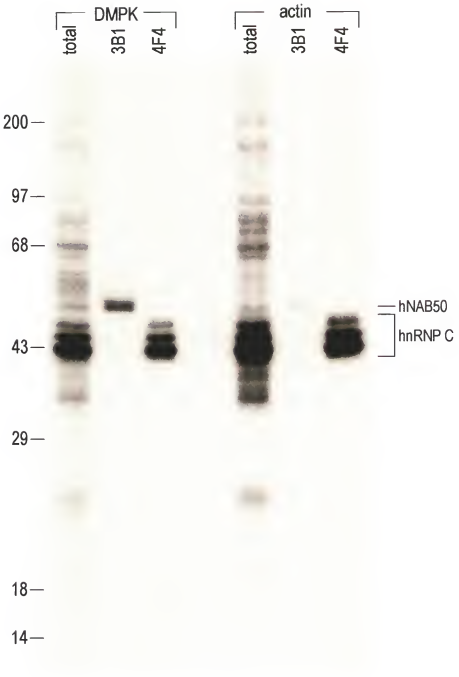
Thus, CUG RNA-binding activity was altered in DM patient cell lines suggesting involvement in the disease process.

Since the bandshifting activity was altered in nuclear fractions in DM cells, we wanted to investigate if there was any alteration in the total protein concentration of hNab50. Immunoblot analysis was performed on whole cell extracts derived from normal and DM patient lymphoblast and myoblast cell lines (Figure 10). As can be seen from this figure, the total amount of hNab50 protein is not significantly different between the different cell lines. Thus, the differences that were seen in CUG binding activity may have been due to differential phosphorylation or some other post-translational modification that has altered hNab50 localization and binding to the $(CUG)_8$ probe.

hNab50 is a DMPK Transcript-Binding Protein

The results described above demonstrated that the hNab50 protein is an hnRNP that binds (CUG) repeats, and poly(A)$^+$ RNA *in vivo*, but is not a stable component of the major hnRNP complex. The structural similarity of hNab50 to the ELR family of proteins suggested that hNab50 may show a binding preference for particular transcripts. To test whether hNab50 associates with DMPK mRNAs, labeled RNAs were prepared by *in vitro* run-off transcription of clones containing either the 3'UTR of DMPK (MTPK.2) or γ-actin genes as a control (Figure 11). The γ-actin gene was chosen because it is relatively U-rich but does not contain any stretches of CUG repeats. These uniformly labeled RNAs were incubated in the presence of HeLa cell nuclear extracts and exposed to UV light to generate covalent crosslinks between proteins and RNA. After treatment with RNase, samples were subjected to immunoprecipitation using MAb 3B1 (anti-

Figure 11. The hNab50 protein in HeLa cell nuclear extracts preferentially photocrosslinks to RNAs containing the 3'-UTR of DMPK. Labeled RNAs containing DMPK and actin 3' -UTR sequences were synthesized in vitro, incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total or immunopurified proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Immunopurifications were performed using either an anti-hnRNP C MAb (4F4) or an anti-hNab50 MAb (3B1). Sizes are indicated in kilodaltons.

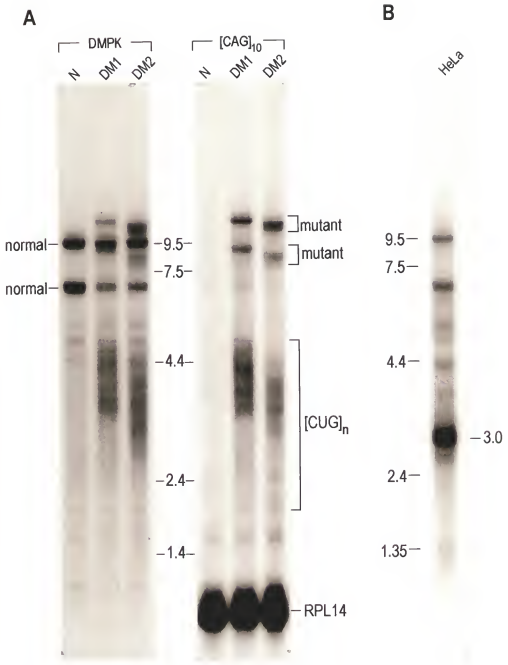

DMPK
total
3B1
4F4
actin
total
3B1
4F4
200
97
68
43
29
18
14
hNAB50
hnRNP C

hNab50) or with MAb 4F4 (anti-hnRNP C) as a control. Samples were fractionated by SDS-PAGE and visualized by autoradiography. Most of the proteins in HeLa nuclear extracts, including hnRNP C, crosslinked more efficiently to actin RNA than to DMPK RNA. In contrast, hNab50 preferentially crosslinked to DMPK RNA. These data are consistent with the idea that hNab50 binds only to a subset of mRNAs and that it may be a transcript-specific binding protein. We speculated that hNab50 may be responsible for post-transcriptional processing of particular genes, a function that has been proposed for several of the ELR proteins due to their transcript binding preferences and functional characterization (Koushika et al., 1996; Palliard et al., 1998; Jain et al., 1997; Chung et al., 1997). Triplet repeat expansion within the 3'-UTR of DMPK could have several effects depending on the level of transcription, stability and location of the enlarged RNA transcript. If the transcript is made and is stable, it could act as an abnormal binding site for hNab50 and other proteins, which could result in aberrant processing/export of DMPK and possibly other pre-mRNAs. We next wanted to investigate the expression of the mutant DMPK mRNA in cell lines derived from myotonic dystrophy patients.

Accumulation of Mutant Transcripts in DM Cell Lines

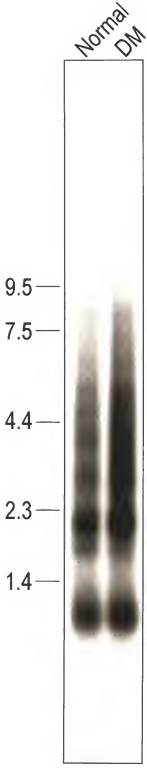
Several investigators have documented the expression of mutant mRNA transcripts in DM patient cells at levels varying from 40 - 70% of wild-type (Taneja et al., 1995; Davis et al., 1997; Hamshere et al., 1997). The defect in DM occurs in the 3'-UTR of the DMPK mRNA, a region of the RNA transcript that is known to be important for 3' end formation, stability and translation (Wahle and Kühn, 1997; Decker and Parker, 1994). We therefore wanted to compare steady-state levels of the DMPK mRNA between normal and DM patient cells. Poly(A)$^+$ RNA

Figure 12. Accumulation of $(CUG)_n$-containing poly(A)$^+$ RNAs in DM lymphoblasts. **(A)** Poly(A)$^+$ RNAs were isolated from either normal (N) or two different DM lymphoblast cell lines (DM1, DM2), fractionated by agarose gel electrophoresis and hybridized with either a DMPK cDNA probe or a $(CAG)_{10}$ oligonucleotide probe. The positions of several different poly(A)$^+$ RNAs are indicated, and include the two normal mRNAs, two mutant DM1 and two mutant DM2 RNAs, the $(CUG)_n$ containing smears in DM1 and DM2, and the 0.96 kb RPL14 mRNA. **(B)** HeLa cell poly(A)$^+$ RNA hybridized with a DMPK specific probe. The major 3.0 kb band is indicated as well as the minor 7.0 and 9.5 kb transcripts.


A
DMPK
N
DM1
DM2
[CAG]10
N
DM1
DM2
normal
normal
9.5
7.5
4.4
2.4
1.4
mutant
mutant
[CUG]n
RPL14
B
HeLa
9.5
7.5
4.4
2.4
1.35
3.0

was isolated from DM and normal patient lymphoblasts by oligo(dT) chromatography and fractionated on a 1% glyoxal agarose gel. In normal lymphoblasts, two major RNA species were visualized at 7.0 and 9.5 kb (Figure 12A). The major reported DMPK transcript in HeLa cells, myoblasts and fibroblasts is 3.0- 3.3 kb although the larger species are present in HeLa cells (Figure 12B and Sabourin et al., 1993). Both the 7.0 and 9.5 kb DMPK RNA species were seen in DM patient lymphoblasts but at reduced levels. In addition, two larger species were seen to migrate ~1.5 kb above the normal bands. The (CUG) repeat expansion in these two patients is approximately 500 repeats, which would result in an increase in the transcript length of 1500 bp. These data strongly suggested that the enlarged transcripts that were seen in the DM patient lines contained the (CUG) repeat expansion. Interestingly, in addition to the enlarged transcripts seen in the DM patient lines, smaller transcripts ranging in size between ~3 kb to 5 kb were also detectable. This suggested that the cell may have difficulty degrading these transcripts and that partially degraded species may accumulate in DM cells. To verify that the enlarged transcripts seen in the DM patient lines contained (CUG) repeats, the RNA blots were stripped and reprobed with an end-labeled $(CAG)_{10}$ oligonucleotide probe which should hybridize to RNAs containing (CUG) repeats. The RNAs detected with this probe co-migrated with the enlarged transcripts that were seen with the DMPK specific probe. The smaller 3-5 kb transcripts were also detected with the $(CAG)_{10}$ probe verifying that they also contained (CUG) repeats. The normal 7.0 and 9.5 kb transcripts in both the DM and normal cell lines were not seen with the (CAG) probe at this exposure level but were detectable with longer exposures. As an additional control, blots were also hybridized with a $(CTG)_{10}$ oligonucleotide probe to ensure that the signal that we were seeing with the $(CAG)_{10}$ probe was indeed due to RNA and not to contaminating DNA (Figure

Figure 13. No change in $(CAG)_n$ containing transcripts in DM patient cells. Poly(A)$^+$ RNAs were isolated from either normal (N) or DM lymphoblast cell lines, fractionated by agarose gel electrophoresis and hybridized with a $(CTG)_{10}$ oligonucleotide probe. Many different $(CAG)_n$ containing transcripts are detected. No transcripts of similar size to DMPK normal or mutant transcripts were detected with this probe.


Normal
DM
9.5
7.5
4.4
2.3
1.4

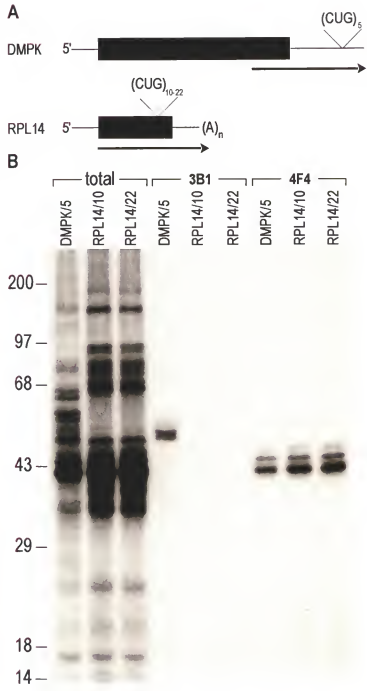
13). The $(CTG)_{10}$ probe did not detect the same bands as the $(CAG)_{10}$ probe, indicating that the signal was not the result of DNA contamination.

The $(CAG)_{10}$ probe also revealed another (CUG) containing transcript that migrated at ~0.9 kb and was present in both DM and normal cell lines (Figure 12A). This RNA was isolated by screening a lymphoblast cDNA library using the $(CAG)_{54}$ probe (Paliouris, 1998) and was revealed to be an mRNA for a ribosomal protein RPL14 (Chan et al., 1996).

Another (CUG) Repeat-Containing mRNA is Not Bound by hNab50

RNA blot analysis described above revealed another (CUG) containing transcript, which was much more abundant than DMPK transcripts. This 0.9 kb mRNA codes for the ribosomal protein RPL14 and contains a polymorphic (CUG) repeat stretch located within the coding region of the protein (Figure 14A) (Chan et al., 1996; Aoki et al., 1996). If hNab50 binding was solely (CUG)-dependent, then the more abundant RPL14 mRNA would out-compete the mutant DMPK mRNAs for binding. To test whether hNab50 was able to bind to the RPL14 mRNA, two RPL14 clones (containing 10 and 22 repeats) together with a clone containing DMPK 3'UTR sequences were transcribed, crosslinked to HeLa nuclear extracts and immunoprecipitated as described above (Figure 14B). While the hnRNP C associated with all three transcripts, hNab50 crosslinked only to the DMPK RNA and not to the RPL14 RNAs. These data not only supported the idea that hNab50 is a transcript-specific binding protein, but also suggested that the context of the (CUG) repeat is important. Either hNab50 does not recognize (CUG) repeats in the context of the RPL14 mRNA or other proteins in HeLa nuclear extracts compete for binding sites. Although crosslinking of hNab50 to the RPL14 RNA was not detected in HeLa nuclear extracts, recombinant hNab50 does crosslink to RPL14 (L. Timchenko,

Figure 14. The hNab50 protein does not crosslink to another (CUG) containing RNA transcript. **(A)** Structure of the DMPK and RPL14 mRNAs. The location of the CUG repeat in the transcript is indicated. The black box indicates the protein coding region while the thin lines represent either 5' or 3'-UTR sequences. The probes used in the crosslinking experiment are indicated by arrows. **(B)** hNab50 crosslinks to the DMPK transcript but fails to crosslink to either RPL14 clone. Labeled RNAs containing DMPK 3'-UTR sequence or RPL14 sequence as indicated in (A) were synthesized in vitro, incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total or immunoprecipitated proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Immunopurifications were performed using either an anti-hnRNP C MAb (4F4) or an anti-hNab50 MAb (3B1). Sizes are indicated in kilodaltons.


A
DMPK
5'
(CUG)5
(A)n
(CUG)10-22
RPL14
5'
(A)n
B
total
3B1
4F4
DMPK/5
RPL14/10
RPL14/22
DMPK/5
RPL14/10
RPL14/22
DMPK/5
RPL14/10
RPL14/22
200
97
68
43
29
18
14

personal communication). This indicated that hNab50 has the ability to bind to the RPL14 substrate at higher concentrations. However, crosslinking was not seen in nuclear extracts because other proteins competed for binding of RPL14 or the affinity of hNab50 for this transcript was low enough that the concentration of hNab50 in nuclear extracts was too low to detect an interaction.

hNab50 Shows Increased Binding to Mutant DMPK Transcripts

In our original hypothesis of DM pathogenesis, we suggested that the enlarged CUG repeat within the 3'UTR of the DMPK transcript acts as a binding site for an RNA-binding protein (Timchenko et al., 1996b; Caskey et al., 1996). The larger the expansion, the more of this protein should bind which, at some level should result in the sequestration of this protein away from other mRNAs. To test whether hNab50 fit this RNA-binding sequestration model, we analyzed crosslinking of mutant DMPK transcripts to hNab50 and compared them to normal DMPK transcripts. Clones containing the 3'UTR of DMPK with variable numbers of (CUG) repeats (0, 6, 54, and 90) were generated using PCR mutagenesis (Figure 15). Uniformly labeled RNAs were transcribed and used in an *in vitro* crosslinking/label transfer experiment with HeLa cell nuclear extracts as described above (Figure 16). hNab50 crosslinked to DMPK 3'-UTR RNA with no repeats, but with reduced efficiency as compared to a normal 3'-UTR containing 6 repeats. Crosslinking of hNab50 increased modestly with larger repeats, but was not proportional to the increase in repeat size. In other words, a ten-fold increase in repeat number did not result in a ten-fold increase in hNab50 crosslinking. Similar results were obtained using different concentrations of substrate RNA (10 fmoles and 50 fmoles of labeled RNA) arguing against a protein titration effect (data not shown). These results suggested that although hNab50

Figure 15. Mutant DMPK 3' UTR constructs containing variable numbers of (CUG) repeats. DMPK 3'-UTR plasmid constructs were generated by PCR mutagenesis to contain no repeats, a normal number of repeats (6), or a mutant number of repeats, (54) and (90). The white region of the box represents DMPK sequence and the black box represents the (CUG) repeats. All constructs were cloned with identical surrounding sequence behind a T7 promoter, as indicated by the bent arrow.

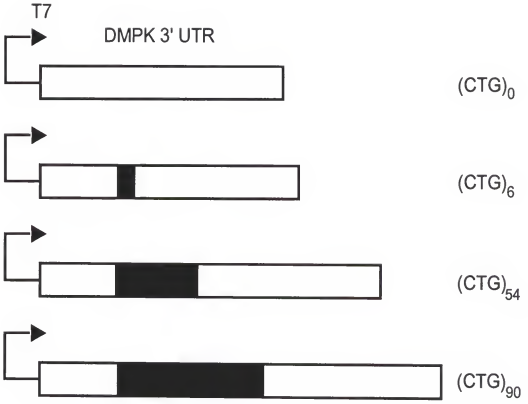
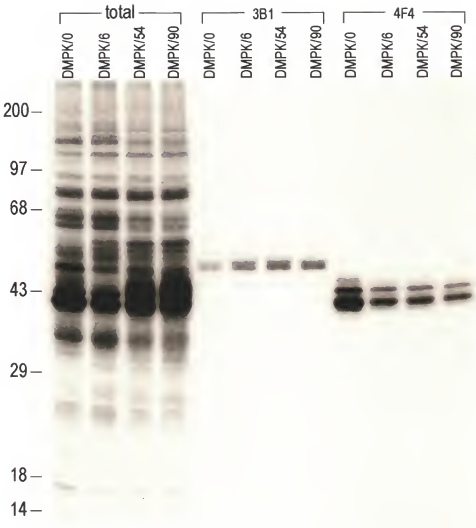

T7
DMPK 3' UTR
(CTG)0
(CTG)6
(CTG)54
(CTG)90

Figure 16. hNab50 shows increased crosslinking to mutant constructs, but the increase is not proportional to repeat size. DMPK 3'-UTR transcripts containing 0, 6, 54 or 90 (CUG) repeats were incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total or immunopurified proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Immunopurifications were performed using either an anti-hnRNP C MAb (4F4) or an anti-hNab50 MAb (3B1). Sizes are indicated in kilodaltons.


total
3B1
4F4
DMPK/0
DMPK/6
DMPK/54
DMPK/90
DMPK/0
DMPK/6
DMPK/54
DMPK/90
DMPK/0
DMPK/6
DMPK/54
DMPK/90
200
97
68
43
29
18
14

binds (CUG) repeats, its association with DMPK 3'-UTR RNAs is modulated, but not dependent, on the triplet repeats contained in the transcript. These results may also be a reflection of the crosslinking efficiency of (CUG) repeats as compared to other sequences within the DMPK 3'UTR or they may result from competition between hNab50 and other (CUG) binding proteins in nuclear extracts. We concluded from these experiments that hNab50 is probably not a simple (CUG) binding protein, but may be a transcript-specific binding protein that is either recruited or stabilized by the presence of the (CUG) repeat and that the (CUG) repeat is part of a cis-acting motif.

To summarize, hNab50 is a novel human hnRNP that is primarily nuclear in localization and associates with poly(A)$^+$ RNA in vivo. It binds a $(CUG)_8$ RNA and associates with the DMPK mRNA in vitro. It is structurally similar to a class of proteins, ELR, which are involved in many aspects of mRNA metabolism such as alternative splicing and mRNA 3'-end formation. Its closest homolog, EDEN-BP, is involved in poly (A) tail length control in *Xenopus* oocytes. The interaction of hNab50 with Nab2p in the two-hybrid system also suggested a role in mRNA 3'-end formation and mRNA export.

Identification of Triplet Repeat Expansion (EXP) RNA-Binding Proteins

To investigate whether additional (CUG) binding proteins existed, the constructs shown in Figure 15 were utilized. Total crosslinked material as shown in Figure 16 was carefully analyzed for the presence of other (CUG) binding proteins. Closer inspection of the total crosslinked proteins at a lower exposure level revealed the expansion (EXP) binding proteins (Figure 17). These 40 – 45 kD proteins did not crosslink detectably to 3'-UTR RNAs containing 6 repeats, or to antisense (as) RNAs but photocrosslinked to DMPK RNA containing 54 or 90

Figure 17. Crosslinking of mutant DMPK transcripts reveals the novel $(CUG)_n$ expansion binding proteins. A normal DMPK transcript containing 6 repeats, an antisense or two mutant transcripts containing 54 and 90 repeats were incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Sizes are indicated in kilodaltons. EXP proteins are indicated at 40 – 45 kD.

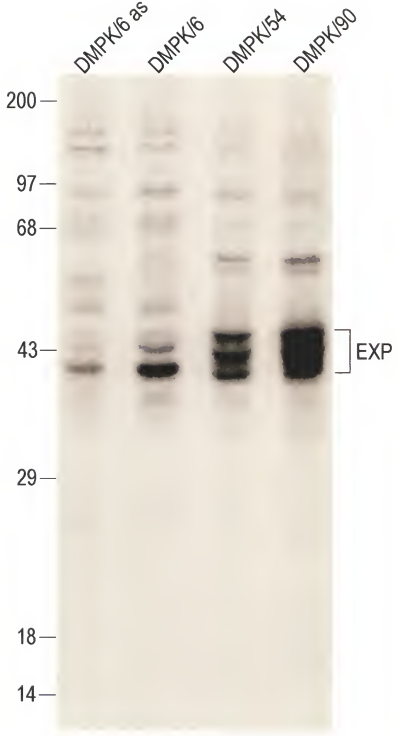
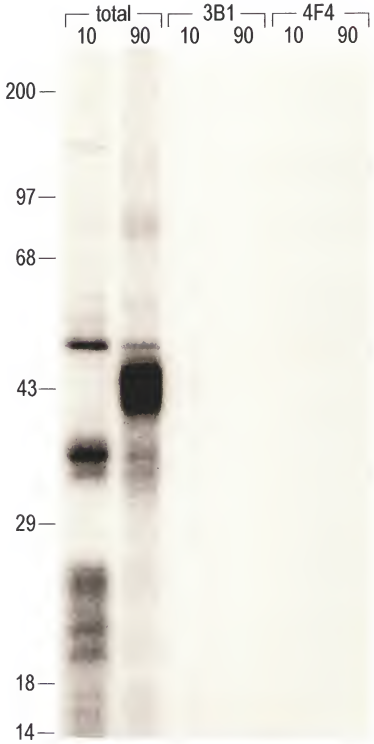

DMPK/6 as
DMPK/6
DMPK/54
DMPK/90
200
97
68
43
29
18
14
EXP

Figure 18. EXP proteins crosslink only to expanded repeats. (CTG) repeats of 10 and 90 were recloned in the absence of DMPK sequence and labeled transcripts were prepared by in vitro transcription using T7 RNA polymerase. The labeled RNAs were incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total or immunopurified proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Immunopurifications were performed using either an anti-hnRNP C MAb (4F4) or an anti-hNab50 MAb (3B1). Neither hNab50 nor hnRNP C are responsible for the EXP activity. Sizes are indicated in kilodaltons.


total
10
90
3B1
10
90
4F4
10
90
200
97
68
43
29
18
14

repeats. To determine if the EXP proteins were simply (CUG)-binding proteins or if they were dependent on DMPK sequence, (CUG) repeats of 10, 54, and 90 were re-cloned in the absence of DMPK transcript sequences and tested by *in vitro* crosslinking and label transfer (Figure 18). The EXP proteins crosslinked to $(CUG)_{54}$ (not shown) and $(CUG)_{90}$ but not to the $(CUG)_{10}$. To determine if hNab50, or hnRNP C proteins were responsible for the EXP activity, immunoprecipitations using MAb 3B1 (anti-hNab50) and 4F4 (anti-hnRNP C) were performed on the crosslinked material. Neither hNab50 nor hnRNP C displayed detectable crosslinking to (CUG) repeats of either construct.

The fact that the EXP proteins crosslinked to (CUG) repeats of 54 and 90, but not to repeats of 10, indicated that these proteins recognized an RNA secondary structure that is not formed with small repeats. Structural studies of (CUG) repeats indicate that they form hairpin structures *in vitro* and that these hairpins are particularly stable with repeats >20 (Figure 19) (Napierala and Krzyosiak, 1997). To further examine the association of EXP proteins with (CUG) repeats, an additional set of clones containing repeat sizes of 10, 20, 35, 74, and 97 were tested (Figure 20). If the EXP proteins recognize double-stranded stem structures formed by these RNAs, the interaction would not occur efficiently with repeat numbers of 10 or less. (CUG) repeats of 20 or more should form stable hairpins and allow binding of the EXP proteins. As the number of repeats increases, the double-stranded portion of the RNA would become larger and more stable and one would expect to see a proportional increase in the binding of the EXP proteins. The results shown in Figure 21A supported the prediction that the EXP proteins would only crosslink to (CUG) repeats of greater than 20. The EXP proteins exhibited a proportional increase in crosslinking activity between $(CUG)_{20}$ and $(CUG)_{97}$. Figure 21B

Figure 19. (CUG) repeats form metastable hairpins. This model depicts the predicted structures of CUG repeat containing RNAs of different lengths based on chemical probing and mutagenesis studies (Napierala and Krzyzosiak, 1997). Repeats of less than 10 form single-stranded structures under the conditions tested. (CUG) repeats of 10 form hairpins that are very unstable and exist in both single and double-stranded forms. (CUG) repeats of 20 or more form stable hairpin structures and may provide a binding site for double-stranded RNA binding proteins on the stem region.

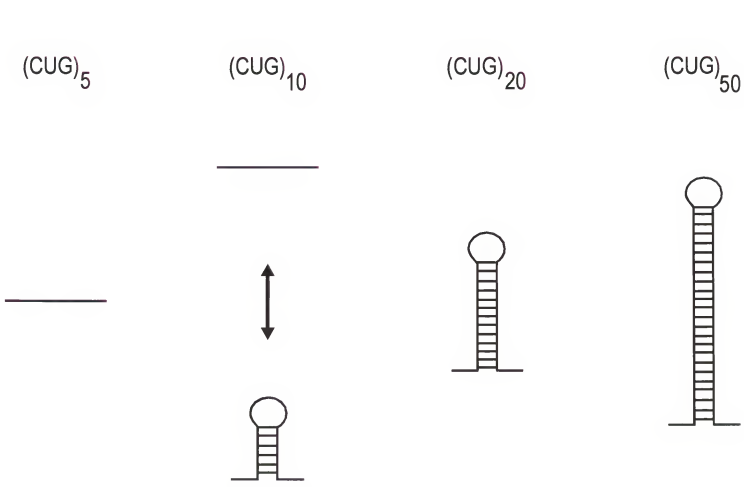

(CUG)5
(CUG)10
(CUG)20
(CUG)50

Figure 20. Pure (CUG) repeat constructs. Pure (CUG) repeats of 11, 20, 35, 74, and 97 were kindly provided by Dr. Charles Thornton (University of Rochester). All are cloned behind a T7 promoter.

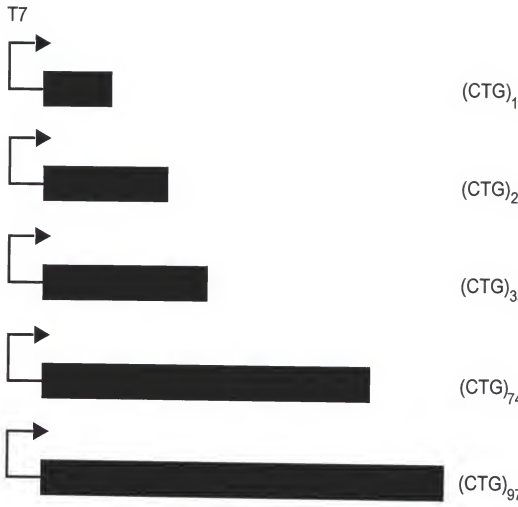
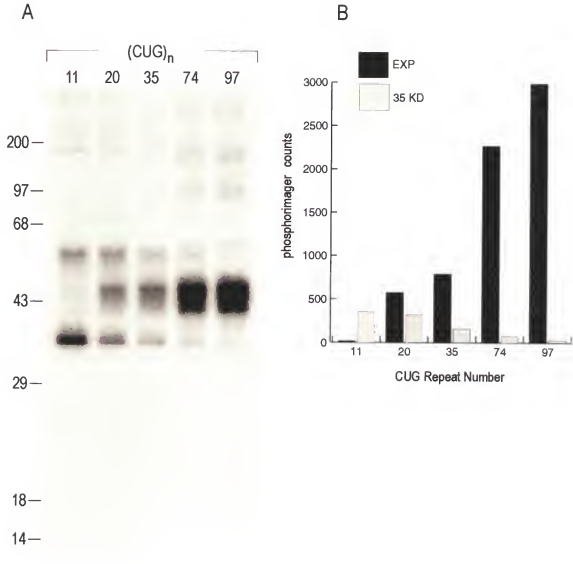

T7
(CTG)10
(CTG)20
(CTG)35
(CTG)74
(CTG)97

Figure 21. Crosslinking of EXP proteins is proportional to repeat size. **(A)** Plasmid constructs containing (CTG) repeats of 11, 20, 35, 74, and 97 were transcribed in vitro using T7 RNA polymerase. The labeled RNAs were incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Sizes are indicated in kilodaltons. **(B)** Crosslinking activity was determined using a phosphorimager and depicted graphically. An unknown 35 kD protein is also indicated as an internal control since the crosslinking of this protein declines with increased $(CUG)_n$ size.


A
(CUG)n
11
20
35
74
97
200
97
68
43
29
18
14
B
EXP
35 KD
phosphorimager counts
3000
2500
2000
1500
1000
500
0
11
20
35
74
97
CUG Repeat Number

compares the EXP protein crosslinking to an unidentified 35 kD protein that shows a progressive decrease in crosslinking activity.

To control for any residual RNA present in the gel and to insure that the signal seen was due only to crosslinked protein, the following controls were performed using $(CUG)_{35}$ and $(CUG)_{74}$ RNAs (Figure 22). RNAs were incubated in nuclear extracts, crosslinked and then digested with RNase A alone or with RNase A and micrococcal nuclease (MN), which will digest both single- and double-stranded RNA. In addition, RNAs were incubated with extracts and then digested with or without crosslinking, and treated with RNase A. Additionally, one sample was digested with proteinase K after crosslinking with the expectation that no signal should be seen if a protein is responsible for the activity. This satisfied us that the activites seen were due to crosslinked protein and not due to undigested RNA.

Since we hypothesized that the EXP proteins bind to double-stranded RNA generated by enlarged (CUG) triplet repeats, we wanted to test whether these proteins could bind to other triplet repeats and double-stranded RNA structures (Figure 23). (CUG) and (CAG) repeats of both 10 or 54 were generated by in vitro transcription and uniformly labeled with $[\alpha^{32}P]$-GTP and crosslinked. The EXP proteins crosslinked to $(CUG)_{54}$, but did not crosslink to (CAG) repeats of either size (Figure 23A). To test a well studied double-stranded RNA hairpin structure, the HIV-1 TAR element was tested for its ability to crosslink to the EXP proteins and to compete for $(CUG)_{90}$ crosslinking (Figure 23B). TAR element forms a hairpin loop with a double-stranded stem and is recognized by both Rev, a double-stranded HIV-1 protein, and TRBP, a cellular protein that contains double-stranded RNA binding motifs (Cullen and Malim, 1991; Gatignol et al., 1991). A second RNA, mTAR, which is a mutant TAR element that is unable to bind TRBP (TAR RNA binding protein) as described previously (Gatignol et al.,

Figure 22. Analysis of EXP protein activity under various conditions. Plasmid constructs containing (CTG) repeats of 35 and 74 were transcribed in vitro using T7 RNA polymerase. The labeled RNAs were incubated in HeLa cell nuclear extracts, and either photocrosslinked with UV light or incubated on ice. Following the crosslinking reactions, the samples were either digested with RNase A, micrococcal nuclease (MN) or with proteinase K (Prot K). Total proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Sizes are indicated in kilodaltons.

crosslink	+	+	−	−	+	+		+	+
RNase A	+	+	+	+	+	+		+	+
MN	+	+	+	+	−	−		+	+
Prot K	−	−	−	−	−	−		+	+

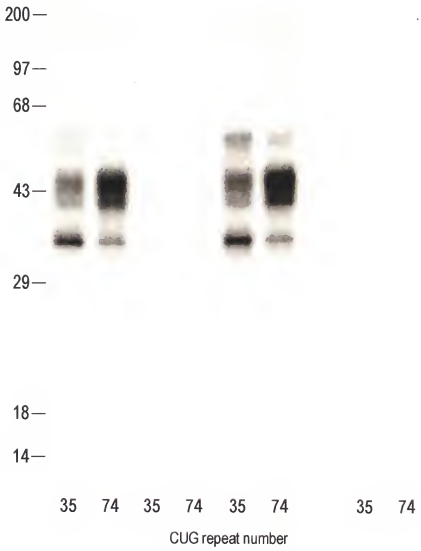
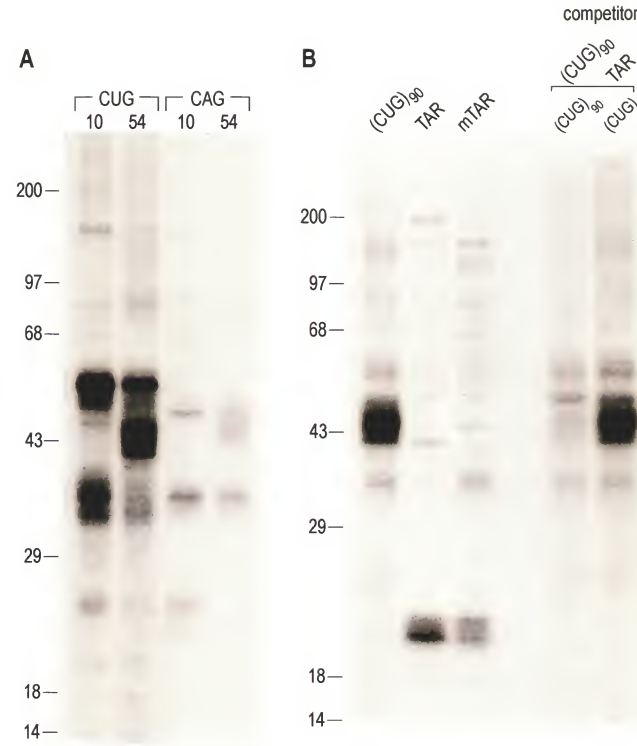


Figure 23. Crosslinking of EXP proteins is specific for (CUG) repeats. **(A)** Plasmid constructs containing (CTG) or (CAG) repeats of 10 or 54 were transcribed in vitro using T7 RNA polymerase in the presence of [α^{32}P]-GTP. The labeled RNAs were incubated in HeLa cell nuclear extracts, and photocrosslinked with UV light. Total proteins were then fractionated by SDS-PAGE, and proteins that bound were detected by label transfer/autoradiography. Sizes are indicated in kilodaltons. **(B)** EXP proteins do not crosslink to the TAR RNA. Plasmids containing either 90 CUG repeats, the TAR RNA sequence (TAR) or a mutant TAR RNA sequence (mTAR) were transcribed with T7 RNA polymerase. The first three lanes represent activity seen with each labeled RNA. In the last two lanes, labeled $(CUG)_{90}$ RNA was mixed with a 500 fold excess of cold $(CUG)_{90}$ or cold TAR RNA prior to addition to the nuclear extract.


A
CUG
CAG
10
54
10
54
200
97
68
43
29
18
14
B
competitor
(CUG)90
TAR
(CUG)90
TAR
mTAR
(CUG)90
(CUG)90
200
97
68
43
29
18
14

1991). Neither the TAR or mTAR crosslinked to the EXP proteins. In addition, competition with >100 fold excess of cold TAR RNA was unable to disrupt EXP crosslinking to a labeled $(CUG)_{90}$ target. Cold $(CUG)_{90}$ showed nearly complete abolishment of labeled crosslinked protein. Thus, not only are the EXP proteins specific for long (CUG) repeats, they show specificity in the presence of other RNA elements with double-stranded structure.

Strategies for Identifying the EXP Proteins

The existence of a protein that shows a high degree of crosslinking to enlarge CUG repeats is very exciting in terms of supporting the sequestration model. While many studies examining RNA metabolism can be accomplished without the identity of these proteins being known, their identification could direct which aspects of cellular metabolism to study. It is possible that the EXP proteins are not involved in mRNA metabolism but rather are part of a completely different cellular process. Therefore, identification of these proteins could potentially provide a major contribution to understanding the pathogenesis of myotonic dystrophy.

Candidate Proteins

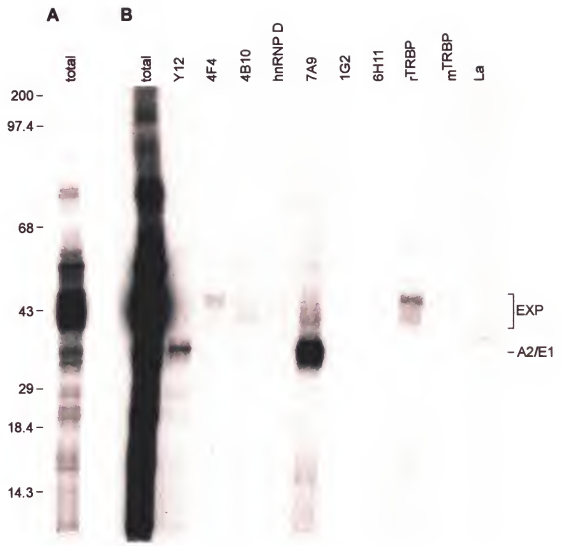
A candidate protein approach was used to try to identify the EXP proteins. This approach was successfully used to identify the first CUG-binding protein, hNab50. The literature was scanned for both single-stranded and double-stranded RNA-binding proteins between 40-50 kD in size for which antibodies are available (see Table 6). HeLa nuclear extracts were crosslinked to a uniformly labeled $(CUG)_{90}$ RNA probe and proteins were immunopurified using each of the antibodies described in Table 6 (See Figure 24). None of the antibodies were able to efficiently immunopurify the EXP proteins in this assay although anti-TRBP antisera did immunopurify a

Table 6

Proteins	Size in kD	Antibody
Nuclear mRNA binding proteins		
hnRNP A1/A2/B1	34, 36, 38	4B10, 7A9[a]
hnRNP C1/C2	41, 43	4F4
hnRNP D	44-48	5B9, rabbit polyclonal
hnRNP E	36-43	7A9
hnRNP G	43	7A9
GRSF	48	1G2, 6H11
Double-stranded RNA-binding proteins		
TRBP	43 and 55	rabbit and mouse polyclonal
La antigen	48	human autoimmune antisera
Ro antigen	52, 60	human autoimmune antisera
Sm proteins	13, 16, 28, 29	Y12
Other		
actin	45	C4

Candidate proteins tested for ability to crosslink to a $(CUG)_{90}$ RNA probe. [a] The 7A9 antibody recognizes several hnRNP proteins.

Figure 24. Immunopurification of EXP proteins using various antibodies against known RNA-binding proteins. **(A)** 24 hour exposure of total crosslinked proteins. **(B)** 72 hour exposure of total crosslinked proteins (total) and crosslinked proteins immunopurified by the various antibodies (see Table 6).


A
B
total
total
Y12
4F4
4B10
hnRNP D
7A9
1G2
6H11
rTRBP
mTRBP
La
200
97.4
68
43
29
18.4
14.3
EXP
A2/E1

faint signal with a long exposure (Figure 24B). Available antibodies against TRBP have not given consistent results by western analysis and must be used at high concentrations to obtain an adequate signal. This may reflect the conservation of this protein between rabbit, mouse and man and the difficulty in breaking immunological tolerance or the lack of immunogenicity of the TRBP protein. The immunopurification of the EXP proteins by anti-TRBP antibodies could result from increased background with this particular antibody or it could be a true, but weak, signal that results from the use of low affinity anti-TRBP antibodies. The fact that cold TAR RNA could not compete with the $(CUG)_{90}$ repeat for EXP binding would argue against the EXP proteins being TRBP. However, TRBP may only bind to the TAR element because of its double-stranded nature but it may have a higher affinity for (CUG) repeats. We are in the process of generating our own MAb directed against TRBP in immunocompromised NZB mice.

Expression Screening

Another approach to isolating the EXP proteins was to directly isolate proteins that could bind a $(CUG)_{90}$ probe in a HeLa λgt11 expression screen. Several other RNA-binding proteins have been successfully isolated using this method, including both La and TRBP which both bind to the TAR RNA element (Gatignol et al., 1991). Unfortunately, although more than a million plaques were screened and different hybridization conditions tried, no $(CUG)_{90}$ RNA-binding proteins were isolated by this method. To insure that the RNA probe was not degraded during the process of hybridization, a sample of the hybridization solution was taken after the incubation period was complete. The sample was extracted and analyzed on a denaturing polyacrylamide gel and was found to be completely intact (data not shown). EXP binding to CUG repeats may involve post-translational modifications that cannot be replicated in *E.coli* or EXP binding may

require a cofactor for binding. We had evidence that ATP was required for crosslinking of EXP proteins to CUG repeat RNAs and that crosslinking was temperature dependent. Even when HeLa nuclear extracts were pre-incubated at 30° C with ATP prior to cooling to 0° C, no crosslinking was detected when RNA is added at 0° C. Pre-structuring of the RNA in nuclear extracts in the presence of ATP, followed by extraction and re-incubation without ATP also did not allow for crosslinking of the EXP proteins to their substrate RNA. We conclude that although this method might be useful for isolating certain types of RNA-binding proteins, protein modification, cofactors or binding conditions prevented the use of these methods for the isolation of the EXP proteins.

RNA Affinity Chromatography

To directly purify the EXP proteins, an RNA affinity column method was used. This method has been successfully used to isolate several RNA-binding proteins including EDEN-BP and hnRNP I (Palliard et al., 1997). The column was prepared by binding unlabeled $(CUG)_{90}$ RNA to a CNBr-activated Sepharose matrix. HeLa nuclear extract supplemented with ATP was bound to the matrix and then washed with buffer containing increasing amounts of KCl up to 1 M. A final wash was performed using 6 M guanidine HCL to strip any residual proteins off the column. Each fraction was dialyzed and tested for the ability to crosslink labeled $(CUG)_{90}$ RNA. Unfortunately, none of the fractions had any crosslinking activity. EXP proteins may have been washed off of the column due to loss of a cofactor or dilution of the ATP. It is also possible that the reaction of attaching the RNA to the column disrupted its structure to a degree that it could no longer be recognized by the EXP proteins.

To address the later problem, $(CUG)_{90}$ was uniformly labeled with digoxigenin conjugated UTP (dig-11-UTP) and incubated with HeLa nuclear extracts under the same conditions used in a crosslinking experiment. The RNA was then selected on protein A Sepharose using anti-digoxigenin antibodies. The bound fraction was washed several times and then proteins were denatured and analyzed by SDS-PAGE/Coomassie staining. As a control, transcripts were double labeled with both dig-11-UTP and [α^{32}P]-GTP and were used in an in vitro crosslinking/label transfer experiment. Crosslinking to the EXP proteins was similar between dig-11-UTP labeled RNAs and control RNAs, indicating that the digoxigenin group did not interfere with EXP binding. However, when the RNAs were isolated onto protein A Sepharose beads, no proteins were detected. As was described above, EXP binding may require a cofactor or some other condition that is lost upon washing of the Sepharose beads.

Preparation of Anti-EXP Antibodies

A fourth approach for isolating the EXP proteins involved the large scale crosslinking of unlabeled $(CUG)_{90}$ RNA containing a poly(A) tail (A_{21}). The RNP/RNA complexes were then purified by oligo(dT) cellulose chromatography and injected into mice for the generation of antibodies. This method has been used successfully to isolate many hnRNP proteins in both yeast and metazoans (Dreyfuss et al., 1993; Anderson et al., 1993; Wilson et al., 1994). In addition, since only a small number of proteins crosslink detectably to the $(CUG)_{90}$ RNA probe by label transfer analysis, the EXP proteins being the most abundant, a limited number of antigens were injected into the mice. Once antibody production is induced, an expression library can be screened and the EXP proteins isolated. Reactive antibodies from a test bleed

recognized a protein of 45 kD in HeLa cells (Figure 25). We are currently testing these antibodies with crosslinked material to determine if they recognize the EXP proteins.

Figure 25. Anti-EXP antibodies recognize a 45 kD protein. Depicted is an immunoblot of total cellular proteins from HeLa cells using either 3B1 MAb against hNab50 or a test bleed from a mouse injected with $(CUG)_{90}$ cross-linked material. Molecular weight is indicated in kilodaltons.

DISCUSSION

Myotonic dystrophy is an autosomal dominant neuromuscular disease that results from a $(CTG)_n$ expansion in the 3'-UTR of the DMPK gene. While other dominantly inherited triplet repeat disorders result from the accumulation of an abnormal protein, the expansion in DM is in a non-coding portion of the gene. Understanding the mechanism of disease pathogenesis in DM has been an area of intense study for the past six years. Several models have emerged to explain the molecular defect in DM but they have failed to explain how the DM triplet repeat expansion causes disease. It is our hypothesis that the repeat expansion exerts a dominant effect at the RNA level. The $(CUG)_n$ repeat acts as a binding site for a transacting factor and affects its function. During this project, I have provided evidence in support of this model by isolating and characterizing two different types of $(CUG)_n$ repeat RNA-binding proteins.

RNA Dominant Mutation Model

Support for an RNA dominant mutation model has been mounting for the past three years. In 1995, Taneja et al. reported in situ hybridization evidence that the mutant DMPK gene was transcribed and mutant transcripts accumulated in the nucleus. Several other groups have also documented the production of mutant RNA transcripts (Wang et al., 1995; Krahe et al., 1995; Sabourin et al., 1995; Davis et al., 1997; Hamshere et al., 1997). Although these studies have reported variable levels of the mutant DMPK

transcripts compared to normal, this may depend on the triplet repeat length or the RNA isolation method used. The studies presented in this dissertation are consistent with these previous studies showing that DMPK mutant genes are transcribed. I also provide evidence that there may be a defect in the normal decay of the mutant DMPK transcripts. In addition to the enlarged transcripts seen migrating above the normal mRNAs, RNAs of heterogeneous size, between 3 and 5 kb, which also contain repeat expansions, were visualized. If these mutant RNAs are sequestered in the nucleus, as is suggested by previous studies (Taneja et al.,1995; Davis et al., 1997; Hamshere et al., 1997), they may be unavailable for complete degradation as a result of being complexed with other RNAs or proteins in the nucleus. Alternatively, the enlarged repeat structure itself may inhibit the normal nuclear RNA decay machinery. Investigators that study RNA decay utilize RNA secondary structures, in the form of poly(G) blocks, to slow the 5'→3' decay machinery and allow for visualization of intermediates (Muhlrad et al., 1994).

In addition to the direct data demonstrating mutant DMPK allele expression, indirect evidence also supports a dominant RNA mutation model. First, no cases of DM have been found to result from a mutation within the coding region of the DMPK gene. Second, neither of the DM related disorders, PROMM and DM2, map to the DMPK locus or to chromosome 19. These facts argue strongly against the primary DM phenotype resulting from alterations in DMPK protein levels or from altered expression of surrounding genes. While it is possible that the other loci represent mutations in different factors of the same pathway, the observation that PROMM may exhibit anticipation argues against this disease resulting from a point mutation in a protein. Too few DM2 patients have been examined to determine if anticipation is a feature of this

disease. I would speculate that PROMM results from a $(CTG)_n$ expansion that is expressed as an RNA. A $(CTG)_n$ expansion, rather than another type of triplet repeat, would produce another expanded $(CUG)_n$ RNA that could bind to the same factors proposed to bind in DM. This would make sense in terms of the multi-systemic nature of PROMM (like DM), similar phenotypic features as DM, but a slightly different distribution and relative severity of affected tissues. In other words, PROMM patients get cataracts, have myotonia and weakness, but with a different muscular distribution. Thus, the $(CTG)_n$ expansion is expressed in a slightly different tissue distribution but ultimately has the same effect. Finally, preliminary studies with transgenic mice containing enlarged CTG repeats under the control of a ubiquitously expressed mammalian promoter display a partial DM phenotype (Monckton et al., 1997b)

Isolation of the First Eukaryotic Triplet-Repeat RNA-Binding Protein

Here we report the isolation and characterization of the first eukaryotic triplet repeat RNA-binding protein, hNab50. I originally isolated hNab50 using the two-hybrid system by employing the yeast hnRNP, Nab2p, as bait. The hNab50 protein was classified as an hnRNP because it was primarily nuclear in its localization and was able to bind poly(A)$^+$ RNA in vivo. Although hNab50 is an hnRNP, it did not co-purify with the major hnRNP complex in HeLa cells. This led us to hypothesize that hNab50 is a transcript-specific binding protein that mediates some aspect of mRNA metabolism for a subset of pre-mRNAs. Subsequently, we discovered that hNab50 was a (CUG) repeat RNA-binding protein which made it a candidate for involvement in myotonic dystrophy.

EMSA analysis demonstrated that the hNab50 protein shifted a $(CUG)_8$ RNA probe and, that $(CUG)_8$ binding activity was altered in DM patient cell lines. Specifically, nuclear CUG-BP1 activity decreased while CUG-BP2 activity increased. These differences do not reflect an overall difference in hNab50 protein concentration in these cells as it was shown to be unchanged. Evidence has been provided that this altered binding activity is the result of differential phosphorylation of the hNab50 protein (Roberts et al., 1997). In addition, it was shown that the DM protein kinase both physically interacts with, and phosphorylates, hNab50 in vitro (Roberts et al., 1997). Thus, reduced DMPK protein levels may result in a shift of hNab50 to a hypophosphorylated state. Although differential phosphorylation affects $(CUG)_8$ binding, it may also affect other functions such as protein localization or interaction with other factors.

In addition to binding and shifting a $(CUG)_8$ probe, hNab50 also bound to the DMPK 3'-UTR in vitro. This binding appeared to be transcript-specific since hNab50 was able to bind the DMPK transcript with no $(CUG)_n$ repeats and did not show significant crosslinking to actin 3'-UTR sequences or the $(CUG)_n$-containing transcript, RPL14. This would suggest that the $(CUG)_n$ repeat is part of a larger cis-acting element and that it is important, but not absolutely necessary, for hNab50 binding. Additionally, increasing the number of $(CUG)_n$ repeats in the 3'-UTR increased the amount of hNab50 that crosslinked, but this increase was not proportional to the increase in repeat size. Structural studies of $(CUG)_n$ repeats revealed that repeats >20 form hairpins with a significant portion in a double-stranded stem (Napierala and Kozosiak, 1997). Since hNab50 possesses single-stranded RNA-binding motifs, this protein probably cannot bind

to a double-stranded structure formed by an enlarged repeat. Thus, the small increase seen with mutant DMPK 3'-UTR sequences may have reflected an increase in single-stranded $(CUG)_n$ repeats at the base of the hairpin. This would also explain why there was no further increase in activity between 54 and 90 repeats. Preliminary electron microscopy data support this conclusion by demonstrating binding of hNab50 at the base of a $(CUG)_{130}$ RNA (C. Urbinati, S. Michalowski, and J. Griffith unpublished data).

Although hNab50 was the first triplet repeat RNA-binding protein isolated from eukaryotes, it was not the first one isolated. In prokaryotes, a triplet repeat RNA-binding protein has been characterized that regulates transcriptional attenuation of the *trp* gene in *Bacillus subtilis*. The *trp* RNA-binding attenuation protein (TRAP), when activated by the presence of tryptophan, binds specifically to an RNA secondary structure, the anti-terminator region, in the nascent *trp* operon leader transcript. Binding of TRAP is dependent on the presence of 11 G/UAG triplet repeats within the leader transcript and disrupts the formation of the anti-terminator RNA structure. This promotes the formation of the terminator and results in transcriptional termination upstream of the *trp* structural genes. TRAP also regulates translation of the *trpE* and *trpG* mRNAs by binding repeat containing hairpins and blocking ribosome access (Babitzke et al., 1994; Antson et al., 1995; Babitzke et al., 1995). This is a vivid example of how the binding of a protein can dramatically change the structure of an RNA molecule.

Nascent RNA pol II transcripts are immediately bound by hnRNPs and snRNPs co-transcriptionally. These proteins are believed to be important in structuring the pre-mRNA molecule for subsequent processing steps. Introduction of an enormous triplet repeat hairpin could disrupt the normal binding of pre-mRNA binding proteins and alter

the processing of the transcript. Several investigators have studied different aspects of RNA metabolism of the mutant DMPK transcript. One investigator documented increased levels of the mutant DMPK pre-mRNA as compared to the normal allele with comparative total steady state levels of both transcripts (Krahe et al., 1995). However, no aberrant splice variants were produced when the mutant DMPK transcript was fully spliced (Krahe et al., 1995). Interestingly, Phillips et al., (1998), showed that overexpression of $(CUG)_n$ repeats (1440 and 960 repeats) by transient transfection into muscle cells led to an alternative splicing pattern of cardiac troponin T mRNA. $(CUG)_n$ expression in vitro switched the splicing pattern from the adult form to the embryonic form by inclusion of exon 5 in the final product. Human cardiac troponin T has a muscle-specific splicing enhancer located downstream of exon 5 and this enhancer contains several interspersed CUG repeats. It was also demonstrated by these authors that hNab50 crosslinks to the enhancer sequence and may be involved in its enhancer activity. It is not clear if altered splicing of cardiac troponin T occurs in DM disease.

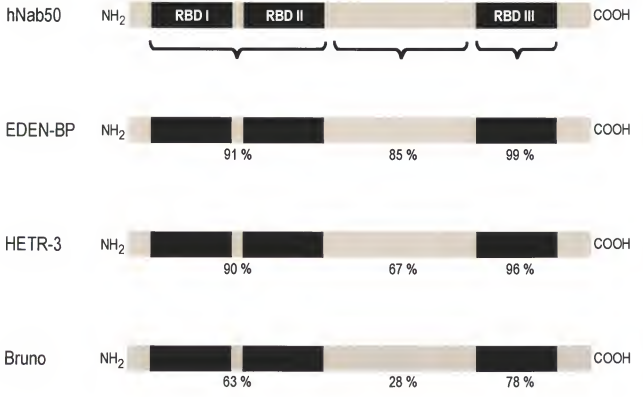
Considering the close proximity of the triplet repeat to the polyadenylation signal in the DMPK gene, it has been suggested that polyadenylation is affected in DM. Wang et al. (1995), provided the first evidence that DMPK mutant transcripts may be hypopolyadenylated. In addition, Hamshere et al., (1997), have shown reduced levels of polyadenylated transcripts in the nuclear fraction. Results to the contrary have also been presented (Davis et al., 1997). Our own data would argue against complete loss of a poly(A) tail, but there is the possibility that DMPK mutant transcripts have shorter than normal tails. This would allow for the isolation on oligo(dT) cellulose, which can bind poly(A) tracts as small as 15 (A) residues, but could affect subsequent pre-mRNA

processing steps and export of the mRNA from the nucleus. Isolation of hNab50 in the two-hybrid system using Nab2p suggests its involvement in polyadenylation. Additionally, antibodies against hNab50 inhibit polyadenylation of a mutant DMPK transcript, but not of the L3 control transcript, in an in vitro polyadenylation assay (C. Urbinati, unpublished data).

The structural similarity between hNab50 with the *elav*-like (ELR) proteins places it in a class of proteins involved in a variety of aspects of mRNA metabolism (Antic and Keene, 1997). As was described in Figure 4, the ELR proteins contain two closely spaced amino-terminal RBDs and one carboxy-terminal RBD separated by what has been termed a "hinge" region. Protein alignments of several different ELR proteins has revealed that the hinge region is highly divergent between the different proteins but that the corresponding RBDs have higher homology. For example, RBD I in hNab50 is more similar to RBD I in other ELR proteins than to the other RBDs within hNab50. This is particularly true of the third RBD where there is 78% amino acid identity between hNab50 and the *Xenopus Etr-1* protein in this region but only 44% homology over the rest of the protein (see Figure 26; Caskey et al., 1996).

A recent study demonstrated that the third RBD of several of the Hu RNA-binding proteins has an affinity for poly(A) of >80 nucleotides in length (Ma et al., 1997; Abe et al., 1996). It has also been demonstrated that these same proteins bind AU-rich elements (AREs) and, in the case of HuD and HuR, do so with the two amino-terminal RBDs (Ma et al., 1996; Chung et al., 1996; King et al., 1994; Gao and Keene, 1996; Chung et al., 1997). AREs are cis-acting sequences found in the 3'-UTR of many short-lived mRNAs, particularly those encoding growth factors and cytokines. There is a

Figure 26 Homology of different ELR proteins with hNab50. Protein alignments were performed (GeneStream alignment tool) to compare the primary amino acid sequence between hNab50, EDEN-BP, HETR-3, and Bruno. EDEN-BP binds to the embryonic deadenylation element found in the 3'-UTR of several maternal mRNAs in *Xenopus* (Paillard et al., 1997). The human ETR-3 protein was isolated in a screen to identify novel proteins in the cardiovascular system but has not been further characterized (Hwang et al, 1994). Bruno is a *Drosophila* ovarian RNA-binding protein, which has been implicated in the translational repression of *oskar*, a developmentally regulated gene (MacDonald et al., 1997).


hNab50
NH2
RBD I
RBD II
RBD III
COOH
EDEN-BP
NH2
COOH
91 %
85 %
99 %
HETR-3
NH2
COOH
90 %
67 %
96 %
Bruno
NH2
COOH
63 %
28 %
78 %

strong correlation between the presence of these elements and the half-life of the mRNA (Chen and Shyu, 1995). The first step in one of the major cytoplasmic mRNA decay pathways is deadenylation of the transcript (Decker and Parker, 1995). It has been speculated that the Hu proteins stabilize the ARE-containing transcripts, possibly in a transcript-specific manner, and are only released when the poly(A) tail is shortened to less than 80 (A) residues (Ma et al., 1997). Interestingly, recent in vivo studies demonstrate that over-expression of HuR stabilizes certain normally unstable ARE-containing mRNAs in vivo (Levy et al., 1998; Fan and Steitz, 1998; Peng et al., 1998).

The *EDEN-BP* is 88.4% identical to hNab50 overall and is 99% identical within the third RBD (see Figure 4 in RESULTS). *EDEN-BP* was identified as a factor that binds to GU-rich embryonic deadenylation element (EDEN) found in several maternal transcripts (see Figure 27). Binding of *EDEN-BP* promoted the rapid deadenylation of these transcripts upon fertilization of the *Xenopus* oocyte (Paillard et al., 1997). Rapid deadenylation differs from the default pathway both kinetically and by the association of a multimeric complex on a specific subset of maternal mRNAs (Paillard et al., 1996). The authors conclude that *EDEN-BP* is a transcript-specific, 3'-UTR binding protein that mediates deadenylation in *Xenopus* embryos (Paillard et al., 1997).

These data suggest that while the functions of ELR proteins are diverse they may accomplish their ends by similar means. In the case of HuR, it may bind concurrently to both the ARE and poly(A) tail and functionally may affect mRNA turnover (Ma et al., 1997). On the basis my own studies and the results described above, I propose a model in which hNab50 is a DMPK pre-mRNA polyadenylation factor which influences poly(A) tail length. This model suggests that hNab50 binds concurrently to a cis-element

Figure 27 EDEN-BP promotes the deadenylation of EDEN-containing transcripts. **(A)** Listed are three of the EDEN motifs that bind EDEN-BP (Paillard et al., 1997). These elements are typically U rich or G/U rich. **(B)** Summary of findings concerning deadenylation and the presence of the different cis-elements found in *Xenopus* maternal mRNAs. The CPE is the cytoplasmic polyadenylation element. This element along with the nuclear polyadenylation element (AAUAAA) is required for cytoplasmic polyadenylation. The embryonic deadenylation element (EDEN) promotes rapid deadenylation of a subset of maternal transcripts upon fertilization of the *Xenopus* oocyte. The presence or absence of each of these elements ultimately determines the fate of the mRNA. It is the combination of these cis-elements and the action of transcript-specific binding factors that allow for the precise control of expression of these transcripts during embryogenesis (Paillard et al., 1997).

A

EDEN-BP binding sequences

Eg5 mRNA	UAUAUAUGUGUGUCUAUC
Eg2 mRNA	UGUCCUUUUAUAUGUAA
c-mos mRNA	UAUAUGUAUGUGUUGUUUUAUGUGUGUGUGUGUGCU

B

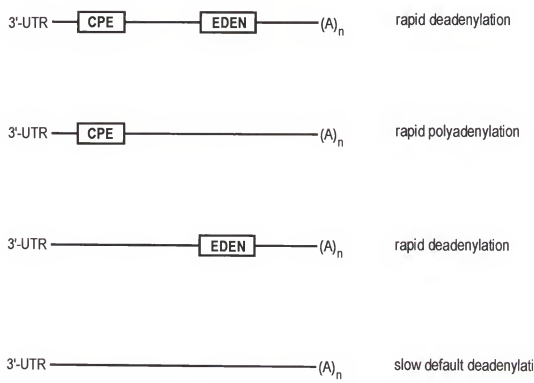
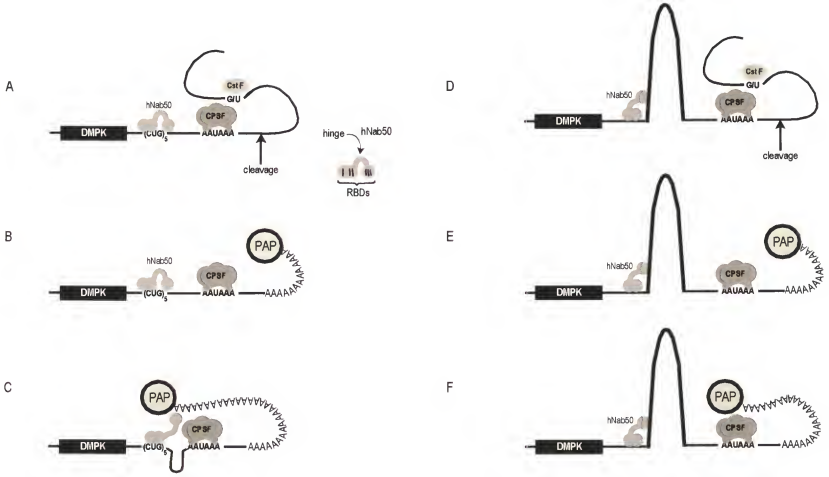


Figure 28. Model for hNab50 function. **(A)** The hNab50 protein associates with the 3'-UTR of the DMPK mRNA but is not involved in the cleavage reaction. In addition to CPSF and CstF, CF I and CF II (not shown) are needed to cleave the transcript **(B)** Distributive polyadenylation is initiated by poly(A) polymerase (PAP). **(C)** The PAB II and hNab50 proteins are involved in the elongation of the poly(A) tail, either in determining the processivity of the reaction or the final length of the tail. **(D)** The triplet repeat expansion does not interfere with the cleavage reaction, although hNab50 binding is altered. **(E)** Distributive polyadenylation is initiated by PAP normally. **(F)** The triplet repeat expansion acts to sterically block the association of hNab50 with the poly(A) tail resulting in a hypopolyadenylated transcript.

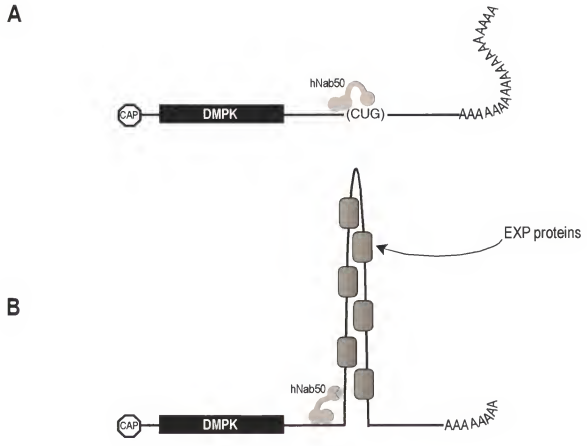

A
CstF
GU
hNab50
CPSF
DMPK
(CUG)n
AAUAAA
cleavage
hinge
hNab50
I II III
RBDs
B
PAP
hNab50
CPSF
DMPK
(CUG)n
AAUAAA
AAAAAAAAAA
C
PAP
CPSF
DMPK
(CUG)n
AAUAAA
AAAAAAAAAA
D
hNab50
CstF
GU
CPSF
DMPK
AAUAAA
cleavage
E
hNab50
PAP
CPSF
DMPK
AAUAAA
AAAAAAAAAA
F
hNab50
PAP
CPSF
DMPK
AAUAAA
AAAAAAAAAA

containing the (CUG) repeat in the DMPK 3'-UTR and the emerging poly(A) tail (Figure 28). Repeat expansion, while allowing more hNab50 to bind, may have the opposite effect on function by sterically hindering association of hNab50 with the poly(A) tail (see Figure 28). This model predicts that expansion results in improper polyadenylation of mutant DMPK transcripts but does not predict how this may impact on DM disease. Although hNab50 did not fit the sequestration model by showing a proportional increase in binding to mutant DMPK transcripts, it did show a slight increase in association with these transcripts. This fact coupled with the nuclear accumulation of mutant transcripts reported in the literature may still result in reduced levels of available hNab50 and could impact on the processing of other transcripts. Additionally, data suggesting differential phosphorylation of hNab50 in DM cells could also alter its function in vivo.

The Expansion (EXP) Binding Proteins

During studies of mutant DMPK transcripts, I discovered the expansion (EXP) RNA-binding proteins. These proteins bind only to enlarged (CUG) repeats >20 and do not depend on DMPK sequence for binding. The EXP activity increases proportionally to the increase in repeat size but the EXP proteins do not bind to expanded (CAG) repeats or to the GC-rich TAR double-stranded RNA hairpin. In addition, competition with >500 fold excess of cold TAR RNA was not able to alter EXP binding to the $(CUG)_{90}$ construct. The EXP proteins fulfill an important prediction of the sequestration model whereby mutant DMPK transcripts act as a molecular sink and sequester the EXP proteins away from their normal cellular function (see Figure 29). Our data suggest that the EXP proteins bind exclusively to the double-stranded region of the repeat expansion

Figure 29. Model depicting the EXP sequestration. **(A)** Normal DMPK mRNA transcript with hNab50 bound. The EXP proteins do not normally bind to this transcript. **(B)** The EXP proteins are sequestered onto the expanded $(CUG)_n$ repeat and are made unavailable for other cellular functions.


A
hNab50
CAP
DMPK
(CUG)
AAAAAAAAAA
B
EXP proteins
hNab50
CAP
DMPK
AAAAAAAA

which would explain why we see a proportional increase in EXP binding. Unlike hNab50, the EXP proteins may not be involved in DMPK pre-mRNA processing since these proteins are not detectably associated with the normal DMPK transcripts.

Although a variety of methods were tried, we have not yet characterized the EXP proteins. Crosslinking of the EXP proteins requires an ATP regeneration system and we speculate that a cofactor may also be required. Preparation of anti-EXP antibodies has resulted in polyclonal sera that recognizes a 45 kD protein by immunoblot analysis (Figure 25), and studies are ongoing to characterize the EXP proteins.

What is the normal cellular function of the EXP proteins? These proteins may or may not be involved in mRNA metabolism. Stable double-stranded RNA structures exist throughout the cell and have a variety of purposes. Ribosomal proteins and RNAs form the ribosome through highly structured and specific interactions between protein and RNA. Spliceosomal snRNPs also possess a large degree of secondary structure and require a host of helicases and other associated factors to form and maintain proper structure (Madhani and Guthrie, 1994). For example, the U4 and U6 snRNAs must form a distinct annealed complex to perform their function in splicing. It was recently demonstrated that the yeast protein, PRP24, facilitates the recycling of U4 and U6 snRNPs following a splicing reaction (Raghunathan and Guthrie, 1998). During the splicing reaction, these snRNAs are separated and must be re-annealed to function in the next reaction. Immunodepletion of PRP24 resulted in the accumulation of unpaired U4 and U6 RNAs making them unavailable for additional splicing reactions. It is conceivable that the EXP proteins are involved in the refolding and maintenance of a specific type of double-stranded RNA structure, which the large (CUG) repeats resemble.

Conclusions and Future Studies

It is our hypothesis that the primary defect in myotonic dystrophy results from the accumulation of mutant DMPK transcripts in cells leading to the alteration of RNA-binding protein function. I have isolated two different types of $(CUG)_n$-binding proteins, hNab50 and the EXP proteins. The hNab50 protein is a single-stranded RNA-binding protein that shows sequence specificity for binding to DMPK transcripts. We predict that hNab50 plays an integral role in processing of a subset of pre-mRNA/mRNAs possibly at the level of polyadenylation. In contrast, the association of the EXP proteins with mutant DMPK transcripts is only induced when the repeat reaches a certain size. We believe that these proteins are double-stranded RNA-binding proteins and that their function is altered in DM as a result of sequestration on mutant DMPK transcripts.

Future work on hNab50 will focus primarily on understanding its role in DMPK pre-mRNA polyadenylation. We have already begun experiments using an in vitro assay system that allows us to test different mutant DMPK constructs. Immunoinhibition studies have already yielded interesting results as discussed above. Immunodepletion of hNab50 and reconstitution with recombinant protein is the next obvious step in understanding its involvement in this process. In vitro polyadenylation assays with purified components will also need to be performed as has been done with the yeast hnRNP, Nab4p (Krecic, 1998). Although CPSF, PAB II, and PAP determine tail length in the adenovirus L3 model RNA, are other factors involved in modulating this effect in a sequence specific manner? Lessons from pre-mRNA splicing, and the recent studies in

the regulation of 3'-end cleavage site selection, would predict that the cell would also carefully regulate poly(A) tail length.

Structural studies are ongoing to crystallize a histidine-tagged version of hNab50 bound to a (CUG) repeat (in collaboration with Y. Shamoo, Rice University). In addition, multi-dimensional NMR spectroscopy of the different RNA-binding domains is being pursued in collaboration with X. Gao (University of Houston). There are still many questions concerning RNA-binding specificity and the elements within DMPK that hNab50 recognizes. His-tagged constructs containing the two amino-terminal RBDs or the carboxy-terminal RBD can be used to narrow down which motifs are important in hNab50 RNA binding. Given the recent data that the carboxy-terminal RBD of HuR binds to long poly(A) tails, it will be interesting to find out if the third RBD in hNab50 also shares this activity. In addition, we will determine the elements in the DMPK 3'-UTR that are required for hNab50 binding and efficient polyadenylation of this transcript. Sequential truncation of the DMPK transcript or hybridization of complimentary oligonucleotides will be useful in determining which elements are important. In addition, movement of the $(CUG)_n$ expansion upstream to a position 5' of the hNab50 binding site could be useful in determining if the expansion exerts its effect by steric hindrance.

Characterization of EXP proteins will continue with the pursuit of anti-EXP antibodies. The preliminary results are promising, and if adequate antibodies are obtained we will isolate the EXP proteins by expression screening. Colocalization with mutant DMPK transcripts within intra-nuclear foci will be studied if high affinity antibodies are obtained. In addition, biochemical isolation of the EXP proteins using crosslinking as an assay will be pursued if anti-EXP antibodies do not prove helpful. It

will be very interesting to determine if either hNab50 or the EXP proteins are involved in DM2 or PROMM. Once EXP antibodies are obtained, subcellular distribution of the EXP proteins in DM2 or PROMM patient cells may yield an answer.

One of the most interesting aspects of this project has been gaining an understanding of the basic processes of mRNA metabolism, and how the cell takes this essential process to a higher level to create the diversity that is needed to operate an entire organism. Understanding how the intricate components of cells exert their control over the entire organism will allow us to better understand the causes of human disease and counteract the misery inflicted on its sufferers. This dissertation describes an attempt to better understand the basic mechanisms of pre-mRNA processing and how a triplet repeat expansion within one mRNA might result in the genetic disease myotonic dystrophy.

REFERENCE

Aaltonen, L.A., P. Peltomaki, F.S. Leach, P. Sistonen, L. Pylkkanen, J.P. Mecklin, H. Jarvinen, S.M. Powell, J. Jen, and S.R. Hamilton 1993. Clues to the pathogenesis of familial colorectal cancer. *Science* **260**: 812-816.

Abbruzzese, C., R. Krahe, M. Liguori, D. Tessarolo, M.J. Siciliano, T. Ashizawa, and M. Giacanelli. 1996. Myotonic dystrophy phenotype without expansion of (CTG)n repeat: an entity distinct from proximal myotonic myopathy (PROMM)? *J Neurol* **243**: 715-721.

Aitchison, J.D., G. Blobel, and M.P. Rout. 1996. Kap104p: a karyopherin involved in the nuclear transport of messenger RNA binding proteins. *Science* **274**: 624-627.

Allain, F.H.T., C.C. Gubser, P.W.A. Howe, K. Nagai, D. Neuhaus, and G. Varani. 1996. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formation. *Nature* **380**: 646-650.

Altschul, S.F., W. Gish, W. Miller, E.W. Myers, and D.J. Lipman. 1990. Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Alzhanova-Ericsson, A.T., X. Sun, N. Visa, E. Kiseleva, T. Wurtz, and B. Daneholt. 1996. A protein of the SR family of splicing factors binds extensively to exonic Balbiani ring pre-mRNA and accompanies the RNA from the gene to the nuclear pore. *Genes Dev* **10**: 2881-2893.

Amero, S.A., G. Raychaudhuri, C.L. Cass, W.J. van Venrooij, W.J. Habets, A.R. Krainer, and A.L. Beyer. 1992. Independent deposition of heterogeneous nuclear ribonucleoproteins and small nuclear ribonucleoprotein particles at sites of transcription. *Proc Natl Acad Sci U S A* **89**: 8409-8413.

Amour, J.A.L., D.G. Monckton, D.L. Neil, K. Tamaki, A. MacLeod, M. Allen, M. Crosier, and A.J. Jeffreys. 1993. Mechanisms of mutation at human minisatellite loci. In *Genome Rearrangement and Stability* (ed. K.E. Davies and S.T. Warren), pp. 43-58. Cold Spring Harbor Laboratory Press, New York.

Anderson, J.T. 1994. Characterization of nuclear polyadenylated RNA-binding proteins from *Saccharomyces cerevisiae*. Ph.D. Dissertation. University of Florida, Gainesville.

Anderson, J.T., S.M. Wilson, K.V. Datar, and M.S. Swanson. 1993. NAB2: a yeast nuclear polyadenylated RNA-binding protein essential for cell viability. *Mol Cell Biol* **13**: 2730-2741.

Antson, A.A., J. Otridge, A.M. Brzozowski, E.J. Dodson, G.G. Dodson, K.S. Wilson, T.M. Smith, M. Yang, T. Kurecki, and P. Gollnick. 1995. The structure of trp RNA-binding attenuation protein. *Nature* **374**: 693-700.

Anvret, M., G. Ahlberg, U. Grandell, B. Hedberg, K. Johnson, and L. Edstrom. 1993. Larger expansions of the CTG repeat in muscle compared to lymphocytes from patients with myotonic dystrophy. *Hum Mol Genet* **2**: 1397-1400.

Aoki, M., L. Koranyi, A.C. Riggs, J. Wasson, K.C. Chiu, M. Vaxillaire, P. Froguel, S. Gough, L. Liu, and H. Donis-Keller. 1996. Identification of trinucleotide repeat-containing genes in human pancreatic islets. *Diabetes* **45**: 157-164.

Ashiya, M. and P.J. Grabowski. 1997. A neuron-specific splicing switch mediated by an array of pre-mRNA repressor sites: evidence of a regulatory role for the polypyrimidine tract binding protein and a brain-specific PTB counterpart. *RNA* **3**: 996-1015.

Babitzke, P. 1997. Regulation of tryptophan biosynthesis: Trp-ing the TRAP or how Bacillus subtilis reinvented the wheel. *Mol Microbiol* **26**: 1-9.

Baeyens, K.J., H.L. De Bondt, and S.R. Holbrook. 1995. Structure of an RNA double helix including uracil-uracil base pairs in an internal loop. *Struct Biol* **2**: 56-62.

Bambara, R.A., R.S. Murante, and L.A. Henricksen. 1997. Enzymes and reactions at the eukaryotic DNA replication fork. *J Biol Chem* **272**: 4647-4650.

Bandzuilis, R., M.S. Swanson, and G. Dreyfuss. 1989. RNA-binding proteins as developmental regulators. *Genes Dev* **3**: 431-437.

Battiste, J.L., H. Mao, N.S. Rao, R. Tan, D.R. Muhandiram, L.E. Kay, A.D. Frankel, and J.R. Williamson. 1996. Alpha helix-RNA major groove recognition in an HIV-1 rev peptide-RRE RNA complex. *Science* **273**: 1547-1551.

Benders, A.A., P.J. Groenen, F.T. Oerlemans, J.H. Veerkamp, and B. Wieringa. 1997. Myotonic dystrophy protein kinase is involved in the modulation of the Ca^{2+} homeostasis in skeletal muscle cells. *J Clin Invest* **100**: 1440-1447.

Beyer, A.L., M.E. Christensen, B.W. Walker, and W.M. LeStourgeon. 1977. Identification and characterization of the packaging proteins of core 40S hnRNP particles. *Cell* **11**: 127-138.

Bhagwati, S., A. Ghatpande, and B. Leung. 1996. Normal levels of DM RNA and myotonin protein kinase in skeletal muscle from adult myotonic dystrophy (DM) patients. *Biochim Biophys Acta* **1317**: 155-157.

Bode, J., T. Schlake, M. Rios-Ramirez, C. Mielke, M. Stengert, V. Kay, and D. Klehr-Wirth. 1995. Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci. *Int Rev Cytol* **162A**: 389-454.

Bogerd, H.P., R.A. Fridell, S. Madore, and B.R. Cullen. 1995. Identification of a novel cellular cofactor for the Rev/Rex class of retroviral regulatory proteins. *Cell* **82**: 485-494.

Boucher, C.A., S.K. King, N. Carey, R. Krahe, C.L. Winchester, S. Rahman, T. Creavin, P. Meghji, M.E. Bailey, F.L. Chartier, S.D. Brown, M.J. Siciliano, and K.J. Johnson. 1995. A novel homeodomain-encoding gene is associated with a large CpG island interrupted by the myotonic dystrophy unstable $(CTG)_n$ repeat. *Hum Mol Genet* **4**: 1919-1925.

Brackenridge, S., H.L. Ashe, M. Giacca, and N.J. Proudfoot. 1997. Transcription and polyadenylation in a short human intergenic region. *Nucleic Acids Res* **25**: 2326-2336.

Brook, J.D., M.E. McCurrach, H.G. Harley, A.J. Buckler, D. Church, H. Aburatani, K. Hunter, V.P. Stanton, J.P. Thirion, T. Hudson, R. Sohn, B. Zemelman, R.G. Snell, S.A. Rundle, S. Crow, J. Davies, P. Shelbourne, J. Buxton, C. Jones, V. Juvonene, K. Johnson, P.S. Harper, D.J. Shaw, and D.E. Housman. 1992. Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* **68**: 799-808.

Brunner, H.G., H.T. Bruggenwirth, W. Nillesen, G. Jansen, B.C. Hamel, R.L. Hoppe, C.E. de Die, C.J. Howeler, B.A. van Oost, B. Wieringa, H.H. Ropers, and H.J.M. Smeets. 1993. Influence of sex of the transmitting parent as well as of parental allele size on the CTG expansion in myotonic dystrophy (DM). *Am J Hum Genet* **53**: 1016-1023.

Burd, C.G. and G. Dreyfuss. 1994. RNA binding specificity of hnRNP A1: Significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* **13**: 1197-1204.

Burd, C.G., M.S. Swanson, M. Gorlach, and G. Dreyfuss. 1989. Primary structures of the heterogeneous nuclear ribonucleoprotein A2, B1, and C2 proteins: a diversity of RNA binding proteins is generated by small peptide inserts. *Proc Natl Acad Sci U S A* **86**: 9788-92.

Campbell, K.H., W.A. Ritchie, and I. Wilmut. 1993. Nuclear-cytoplasmic interactions during the first cell cycle of nuclear transfer reconstructed bovine embryos:

implications for deoxyribonucleic acid replication and development. *Biol Reprod* **49**: 933-942.

Campos, A.R., D. Grossman, and K. White. 1985. Mutant alleles at the locus *elav* in *Drosophila melanogaster* lead to nervous system defects: A developmental-genetic analysis. *J Neurogenet.* **2**: 197-218.

Campuzano, V., L. Montermini, M.D. Molto, L. Pianese, M. Cossee, F. Cavalcanti, E. Monros, F. Rodius, F. Duclos, and A. Monticelli 1996. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science* **271**: 1423-1427.

Carango, P., J.E. Noble, H.G. Marks, and V.L. Funanage. 1993. Absence of myotonic dystrophy protein kinase (DMPK) mRNA as a result of a triplet repeat expansion in myotonic dystrophy. *Genomics* **18**: 340-348.

Caskey, C.T., M.S. Swanson, and L.T. Timchenko. 1996. Myotonic dystrophy: discussion of molecular mechanism. *Cold Spring Harb Symp Quant Biol* **61**: 607-614.

Chan, Y.L., J. Olvera, and I.G. Wool. 1996. The primary structure of rat ribosomal protein L14. *Biochem Biophys Res Commun* **222**: 427-431.

Choi, Y.D. and G. Dreyfuss. 1984. Isolation of the heterogeneous nuclear RNA-ribonucleoprotein complex (hnRNP): a unique supramolecular assembly. *Proc Natl Acad Sci U S A* **81**: 7471-7475.

Choi, Y.D., P.J. Grabowski, P.A. Sharp, and G. Dreyfuss. 1986. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science* **231**: 1534-1539.

Chung, S., M. Eckrich, N. Perrone-Bizzozero, D.T. Kohn, and H. Furneaux. 1997. The Elav-like proteins bind to a conserved regulatory element in the 3'- untranslated region of GAP-43 mRNA. *J Biol Chem* **272**: 6593-6598.

Cobo, A., J.M. Martinez, L. Martorell, M. Baiget, and K. Johnson. 1993. Molecular diagnosis of homozygous myotonic dystrophy in two asymptomatic sisters. *Hum Mol Genet* **2**: 711-715.

Colgan, D.F. and J.L. Manley. 1997. Mechanism and regulation of mRNA polyadenylation. *Genes Dev* **11**: 2755-2766.

Cooper, T.A. and W. Mattox. 1997. The regulation of splice-site selection, and its role in human disease. *Am J Hum Genet* **61**: 259-266.

Cullen, B.R. and M.H. Malim. 1991. The HIV1 Rev protein: prototype of a novel class of eukaryotic post-transcriptional regulators. *Trends Biochem Sci* **16**: 346-350.

Daneholt, B. 1997. A look at messenger RNP moving through the nuclear pore. *Cell* **88**: 585-588.

Datar, K.V., G. Dreyfuss, and M.S. Swanson. 1993. The human hnRNP M proteins: identification of a methionine/arginine-rich repeat motif in ribonucleoproteins. *Nucleic Acids Res* **21**: 439-446.

David, G., N. Abbas, G. Stevanin, A. Durr, G. Yvert, G. Cancel, C. Weber, G. Imbert, F. Saudou, E. Antoniou, H. Drabkin, R. Gemmill, P. Giunti, A. Benomar, N. Wood, M. Ruberg, Y. Agid, J.L. Mandel, and A. Brice. 1997. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* **17**: 65-70.

Davis, B.M., M.E. McCurrach, K.L. Taneja, R.H. Singer, and D.E. Housman. 1997. Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. *Proc Natl Acad Sci U S A* **94**: 7388-7393.

Davis, L.I. 1995. The nuclear pore complex. *Ann Rev Biochem* **64**: 865-896.

de Die-Smulders, C.E., C.J. Howeler, C. Thijs, J.F. Mirandolle, H.B. Anten, H.J. Smeets, K.E. Chandler, and J.P. Geraedts. 1998. Age and causes of death in adult-onset myotonic dystrophy. *Brain* **121**: 1557-1563.

de Die-Smulders, C.E., H.J. Smeets, W. Loots, H.B. Anten, J.F. Mirandolle, J.P. Geraedts, and C.J. Howeler. 1997. Paternal transmission of congenital myotonic dystrophy. *J Med Genet* **34**: 930-933.

Decker, C.J. and R. Parker. 1994. Mechanisms of mRNA degradation in eukaryotes. *Trends Biochem Sci* **19**: 336-340.

Della, N.G., P.V. Senior, and D.D. Bowtell. 1993. Isolation and characterisation of murine homologues of the Drosophila seven in absentia gene (sina). *Development* **117**: 1333-1343.

Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* **11**: 1475-1489.

Dreyfuss, G. 1986. Structure and function of nuclear and cytoplasmic ribonucleoprotein particles. *Ann Rev Cell Biol* **2**: 459-498.

Dreyfuss, G., M.J. Matunis, S. Pinol-Roma, and C.G. Burd. 1993. hnRNP proteins and the biogenesis of mRNA. *Ann Rev Biochem* **62**: 289-321.

Dreyfuss, G., M.S. Swanson, and S. Pinol-Roma. 1988. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. *Trends Biochem Sci* **13**: 85-91.

Dunn, P.W., E.T. Walch, and H.F. Epstein. 1994. Phosphorylation reactions of recombinant human myotonic dystrophy protein kinase and their inhibition. *Biochemistry* **33**: 10809-10814.

Dürr, A., M. Cossee, Y. Agid, V. Campuzano, C. Mignard, C. Penet, J.L. Mandel, A. Brice, and M. Koenig. 1996. Clinical and genetic abnormalities in patients with Friedreich's ataxia. *New Eng J Med* **335**: 1169-1175.

Dworetzky, S.I., R.E. Lanford, and C.M. Feldherr. 1988. The effects of variations in the number and sequence of targeting signals on nuclear uptake. *J Cell Biol* **107**: 1279-1287.

Eckner, R., W. Ellmeier, and M.L. Birnstiel. 1991. Mature mRNA 3' end formation stimulates RNA export from the nucleus. *EMBO J* **10**: 3513-3522.

Edwalds-Gilbert, G., K.L. Veraldi, and C. Milcarek. 1997. Alternative poly(A) site selection in complex transcription units: means to an end? *Nucleic Acids Res* **25**: 2547-2561.

Ernst, R.K., M. Bray, D. Rekosh, and M.L. Hammarskjold. 1997. A structured retroviral RNA element that mediates nucleocytoplasmic export of intron-containing RNA. *Mol Cell Biol* **17**: 135-144.

Fan, X.C. and J.A. Steitz. 1998. Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *EMBO J* **17**: 3448-3460.

Fields, S. and O.-K. Song. 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245-246.

Flaherty, S.M., P. Fortes, E. Izaurralde, I.W. Mattaj, and G.M. Gilmartin. 1997. Participation of the nuclear cap binding complex in pre-mRNA 3' processing. *Proc Natl Acad Sci U S A* **94**: 11893-11898.

Fornerod, M., M. Ohno, M. Yoshida, and I.W. Mattaj. 1997. CRM1 is an export receptor for leucine-rich nuclear export signals. *Cell* **90**: 1051-1060.

Freudenreich, C.H., J.B. Stavenhagen, and V.A. Zakian. 1997. Stability of a CTG/CAG trinucleotide repeat in yeast is dependent on its orientation in the genome. *Mol Cell Biol* **17**: 2090-2098.

Fritz, C.C. and M.R. Green. 1996. HIV Rev uses a conserved cellular protein export pathway for the nucleocytoplasmic transport of viral RNAs. *Curr Biology* **6**: 848-854.

Fu, Y.H., D.L. Friedman, S. Richards, J.A. Pearlman, R.A. Gibbs, A. Pizzuti, T. Ashizawa, M.B. Perryman, G. Scarlato, R.G. Fenwick, Jr., and C.T. Caskey. 1993. Decreased expression of myotonin-protein kinase messenger RNA and protein in adult form of myotonic dystrophy. *Science* **260**: 235-238.

Fu, Y.H., A. Pizzuti, R.G. Fenwick, Jr., J. King, S. Rajnarayan, P.W. Dunne, J. Dubel, G.A. Nasser, T. Ashizawa, P. de Jong, B. Wieringa, R. Korneluk, M.B. Perryman, H.F. Epstein, and C.T. Caskey. 1992. An unstable triplet repeat in a gene related to myotonic muscular dystrophy. *Science* **255**: 1256-1258.

Fu, X.D. 1995. The superfamily of arginine/serine-rich splicing factors. *RNA* **1**: 663-680.

Fu, X.D. and T. Maniatis. 1992. Isolation of a complementary DNA that encodes the mammalian splicing factor SC35. *Science* **256**: 535-538.

Gacy, A.M., G. Goellner, N. Juranic, S. Macura, and C.T. McMurray. 1995. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* **81**: 533-540.

Gatignol, A., A. Buckler-White, B. Berkhout, and K.T. Jeang. 1991. Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. *Science* **251**: 1597-1600.

Gatignol, A., C. Buckler, and K.T. Jeang. 1993. Relatedness of an RNA-binding motif in human immunodeficiency virus type 1 TAR RNA-binding protein TRBP to human P1/dsI kinase and Drosophila Staufen. *Mol Cell Biol* **13**: 2193-2202.

Gecz, J., A.K. Gedeon, G.R. Sutherland, and J.C. Mulley. 1996. Identification of the gene FMR2, associated with FRAXE mental retardation. *Nat Genet* **13**: 105-108.

Gerace, L. 1995. Nuclear export signals and the fast track to the cytoplasm. *Cell* **82**: 341-344.

Ghetti, A., S. Pinol-Roma, W.M. Michael, C. Morandi, and G. Dreyfuss. 1992. hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res* **20**: 3671-3678.

Ghisolfi, L., G. Joseph, F. Amalric, and M. Erard. 1992. The glycine-rich domain of nucleolin has an unusual super-secondary structure responsible for its RNA-helix-destabilizing properties. *J Biol Chem* **267**: 2955-2959.

Gohring, F., B.L. Schwab, P. Nicotera, M. Leist, and F.O. Fackelmayer. 1997. The novel SAR-binding domain of scaffold attachment factor A (SAF-A) is a target in apoptotic nuclear breakdown. *EMBO J* **16**: 7361-7371.

Gordenin, D.A., T.A. Kunkel, and M.A. Resnick. 1997. Repeat expansion--all in a flap? *Nat Genet* **16**: 116-118.

Gorlich, D. and I.W. Mattaj. 1997. Nucleocytoplasmic Transport. *Science* **271**: 1513-1518.

Gourdon, G., F. Radvanyi, A.S. Lia, C. Duros, M. Blanche, M. Abitbol, C. Junien, and H. Hofmann-Radvanyi. 1997. Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. *Nat Genet* **15**: 190-192.

Grabowski, P.J. 1998. Splicing regulation in neurons: tinkering with cell-specific control. *Cell* **92**: 709-712.

Green, M.R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Ann Rev Cell Biol* 7: 559-599.

Greenberg, J.R. 1980. Proteins crosslinked to messenger RNA by irradiating polyribosomes with ultraviolet light. *Nucleic Acids Res* **8**: 5685-5701.

Gu, Y., Y. Shen, R.A. Gibbs, and D.L. Nelson. 1996. Identification of FMR2, a novel gene associated with the FRAXE CCG repeat and CpG island. *Nat Genet* **13**: 109-113.

Hamshere, M.G., E.E. Newman, M. Alwazzan, B.S. Athwal, and J.D. Brook. 1997. Transcriptional abnormality in myotonic dystrophy affects DMPK but not neighboring genes. *Proc Natl Acad Sci U S A* **94**: 7394-7399.

Harper, P.S. 1989. Myotonic Dystrophy. Saunders, London/Philadelphia.

Harper, P.S. and R. Rudel. 1994. Myotonic Dystrophy. In *Myology* (ed. A.G. Engel and C. Franzini-Armstrong), pp. 1192-1219. McGraw-Hill, Inc., New York.

Heath, S.K., S. Carne, C. Hoyle, K.J. Johnson, and D.J. Wells. 1997. Characterisation of expression of mDMAHP, a homeodomain-encoding gene at the murine DM locus. *Hum Mol Genet* **6**: 651-657.

Hofmann-Radvanyi, H., C. Lavedan, J.P. Rabes, D. Savoy, C. Duros, K. Johnson, and C. Junien. 1993. Myotonic dystrophy: absence of the CTG enlarged transcript in congenital forms, and low expression of the normal allele. *Hum Mol Genet* **2**: 1263-1266.

Huang, S., T.J. Deerinck, M.H. Ellisman, and D.L. Spector. 1997. The dynamic organization of the perinucleolar compartment in the cell nucleus. *J Cell Biol* **137**: 965-974.

Hunter, A., C. Tsilfidis, G. Mettler, P. Jacob, M. Mahadevan, L. Surh, and R. Korneluk. 1992. The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. *J Med Genet* **29**: 774-779.

The Huntington's Disease Collaborative Research Group 1993. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **72**: 971-983.

Hwang, D.M., W.-S. Hwang, and C.-C. Liew. 1994. Single pass sequencing of a unidirectional human fetal heart cDNA library to discover novel genes of the cardiovascular system. *J Molec Card* **26**: 1329-1333.

Imbert, G., C. Kretz, K. Johnson, and J.L. Mandel. 1993. Origin of the expansion mutation in myotonic dystrophy. *Nat Genet* **4**: 72-76.

Izaurralde, E., J. Lewis, C. McGuigan, M. Jankowska, E. Darzynkiewicz, and I.W. Mattaj. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78**: 657-668.

Jackson, R.J. and N. Standart. 1990. Do the poly(A) and 3' untranslated region control mRNA translation? *Cell* **62**: 15-24.

Jacobs, A.E., A.A. Benders, A. Oosterhof, J.H. Veerkamp, P. van Mier, R.A. Wevers, and E.M. Joosten. 1990. The calcium homeostasis and the membrane potential of cultured muscle cells from patients with myotonic dystrophy. *Biochim Biophys Acta* **1096**: 14-19.

Jain, R.G., L.G. Andrews, K.M. McGowan, P.H. Pekala, and J.D. Keene. 1997. Ectopic expression of Hel-N1, an RNA-binding protein, increases glucose transporter (GLUT1) expression in 3T3-L1 adipocytes. *Mol Cell Biol* **17**: 954-962.

Jansen, G., D. Bachner, M. Coerwinkel, N. Wormskamp, H. Hameister, and B. Wieringa. 1995. Structural organization and developmental expression pattern of the mouse WD-repeat gene DMR-N9 immediately upstream of the myotonic dystrophy locus. *Hum Mol Genet* **4**: 843-852.

Jansen, G., M. Bartolomei, V. Kalscheuer, G. Merkx, N. Wormskamp, E. Mariman, D. Smeets, H. Ropers, and B. Wieringa. 1993. No imprinting involved in the expression of DM-kinase mRNAs in mouse and human tissues. *Hum Mol Genet* **2**: 1221-1227.

Jansen, G., P.J. Groenen, D. Bachner, P.H. Jap, M. Coerwinkel, F. Oerlemans, W. van den Broek, B. Gohlsch, D. Pette, J.J. Plomp, P.C. Molenaar, M.G. Nederhoff, C.J. van Echteld, M. Dekker, A. Berns, H. Hameister, and B. Wieringa. 1996. Abnormal myotonic dystrophy protein kinase levels produce only mild myopathy in mice. *Nat Genet* **13**: 316-324.

Jansen, G., M. Mahadevan, C. Amemiya, N. Wormskamp, B. Segers, W. Hendriks, K. O'Hoy, S. Baird, L. Sabourin, G. Lennon, P.L. Jap, D. Iles, M. Coerwinkel, M. Hofker, A.V. Carrano, P.J. de Jong, R.G. Korneluk, and B. Wieringa. 1992. Characterization of the myotonic dystrophy region predicts multiple protein isoform-encoding mRNAs. *Nat Genet* **1**: 261-266.

Jansen, G., P. Willems, M. Coerwinkel, W. Nillesen, H. Smeets, L. Vits, C. Howeler, H. Brunner, and B. Wieringa. 1994. Gonosomal mosaicism in myotonic dystrophy patients: involvement of mitotic events in $(CTG)_n$ repeat variation and selection against extreme expansion in sperm. *Am J Hum Genet* **54**: 575-585.

Jaspert, A., R. Fahsold, H. Grehl, and D. Claus. 1995. Myotonic dystrophy: correlation of clinical symptoms with the size of the CTG trinucleotide repeat. *J Neurol* **242**: 99-104.

Joseph, B., M. Orlian, and H. Furneaux. 1998. p21(waf1) mRNA contains a conserved element in its 3'-untranslated region that is bound by the Elav-like mRNA-stabilizing proteins. *J Biol Chem* **273**: 20511-20516.

Justice, R.W., O. Zilian, D.F. Woods, M. Noll, and P.J. Bryant. 1995. The *Drosophila* tumor suppressor gene *warts* encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. *Genes Dev* **9**: 534-546.

Kajita, Y., J. Nakayama, M. Aizawa, and F. Ishikawa. 1995. The UUAG-specific RNA binding protein, heterogeneous nuclear ribonucleoprotein D0. Common modular structure and binding properties of the 2xRBD-Gly family. *J Biol Chem* **270**: 22167-22175.

Kang, S., A. Jaworski, K. Ohshima, and R.D. Wells. 1995. Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in E. coli. *Nat Genet* **10**: 213-218.

Kato, K. 1992. Finding new genes in the nervous system by cDNA analysis. *Trends Neurol Sci* **15**: 319-323.

Kawaguchi, Y., T. Okamoto, M. Taniwaki, M. Aizawa, and M. Inoue 1994. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* **8**: 221-228.

Keller, W. and L. Minvielle-Sebastia. 1997. A comparison of mammalian and yeast pre-mRNA 3'-end processing. *Curr Op Cell Biol* **9**: 329-336.

Kharrat, A., M.J. Macias, T.J. Gibson, M. Nilges, and A. Pastore. 1995. Structure of the dsRNA binding domain of E. coli RNase III. *EMBO J* **14**: 3572-3584.

Kiledjian, M., C.G. Burd, M. Gorlach, D.S. Portman, and G. Dreyfuss. 1994. Structure and function of hnRNP proteins. In *RNA-Protein Interactions* (ed. K. Nagai and I.W. Mattaj), pp. 127 - 149. Oxford University Press, Oxford.

Kiledjian, M., C.T. DeMaria, G. Brewer, and K. Novick. 1997. Identification of AUF1 (heterogeneous nuclear ribonucleoprotein D) as a component of the alpha-globin mRNA stability complex. *Mol Cell Biol* **17**: 4870-4876.

Kinoshita, M., R. Takahashi, T. Hasegawa, T. Komori, R. Nagasawa, K. Hirose, and H. Tanabe. 1996. $(CTG)_n$ expansions in various tissues from a myotonic dystrophy patient. *Muscle and Nerve* **19**: 240-242.

Klesert, T.R., A.D. Otten, T.D. Bird, and S.J. Tapscott. 1997. Trinucleotide repeat expansion at the myotonic dystrophy locus reduces expression of DMAHP. *Nat Genet* **16**: 402-406.

Koide, R., I. Ikeuchi, O. Onodera, H. Tanaka, and S. Igarishi 1994. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat Genet* **6**: 9-13.

Korade-Mirnics, Z., P. Babitzke, and E. Hoffman. 1998. Myotonic dystrophy: molecular windows on a complex etiology. *Nucleic Acids Res* **26**: 1363-1368.

Koushika, S.P., M.J. Lisbin, and K. White. 1996. ELAV, a Drosophila neuron-specific protein, mediates the generation of an alternatively spliced neural protein isoform. *Curr Biol* **6**: 1634-1641.

Krahe, R., T. Ashizawa, C. Abbruzzese, E. Roeder, P. Carango, M. Giacanelli, V.L. Funanage, and M.J. Siciliano. 1995. Effect of myotonic dystrophy trinucleotide repeat expansion on DMPK transcription and processing. *Genomics* **28**: 1-14.

Krämer, A. 1996. The structure and function of protein involved in mammalian pre-mRNA splicing. *Ann Rev Biochem* **65**: 367-409.

Krecic, A.M. 1998. Nuclear and Cytoplasmic Functions of a Yeast Heterogeneous Nuclear Ribonucleoprotein. Ph.D. Dissertation. University of Florida, Gainesville.

Kremer, E.J., M. Pritchard, M. Lynch, S. Yu, and K. Holman 1991. Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence $p(CGG)_n$. *Science* **252**: 1711-1714.

La Spada, A.R., E.M. Wilson, D.B. Lubahn, A.E. Harding, and K.H. Fischbeck. 1991. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* **352**: 77-79.

Lafreniere, R.G., D.L. Rochefort, N. Chretien, J.M. Rommens, J.I. Cochius, R. Kalviainen, U. Nousiainen, G. Patry, k. Farrell, B. Soderfeldt, A. Federico, and B.R. Hale 1997. Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EpM1. *Nat Genet* **15**: 298-302.

Lalioti, M.D., H.S. Scott, C. Buresi, C. Rossier, A. Bottani, M.A. Morris, A. Malafosse, and S.E. Antonarakis. 1997. Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature* **386**: 847-851.

Levy, N.S., S. Chung, H. Furneaux, and A.P. Levy. 1998. Hypoxic stabilization of vascular endothelial growth factor mRNA by the RNA-binding protein HuR. *J Biol Chem* **273**: 6417-6423.

Lewis, J.D. and E. Izaurralde. 1997. The role of the cap structure in RNA processing and nuclear export. *Eur J Biochem* **247**: 461-469.

Lewis, J.D., E. Izaurralde, A. Jarmolowski, C. McGuigan, and I.W. Mattaj. 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev* **10**: 1683-1698.

Lia, A.S., H. Seznec, H. Hofmann-Radvanyi, F. Radvanyi, C. Duros, C. Saquet, M. Blanche, C. Junien, and G. Gourdon. 1998. Somatic instability of the CTG repeat in mice transgenic for the myotonic dystrophy region is age dependent but not correlated to the relative intertissue transcription levels and proliferative capacities. *Hum Mol Genet* **7**: 1285-1291.

Lin, C.-L.G., L.A. Bristol, L. Jin, M. Dykes-Hoberg, T. Crawford, L. Clawson, and J.D. Rothstein. 1998. Aberrant RNA processing in a neurodegenerative disease: the cause for absent EAAT2 , a glutamate transporter, in amyotrophic lateral sclerosis. *Neuron* **20**: 589-602.

Maeda, M., C.S. Taft, E.W. Bush, E. Holder, W.M. Bailey, H. Neville, M.B. Perryman, and R.D. Bies. 1995. Identification, tissue-specific expression, and subcellular localization of the 80- and 71-kDa forms of myotonic dystrophy kinase protein. *J Biol Chem* **270**: 20246-20249.

Mahadevan, M.S., C. Amemiya, G. Jansen, L. Sabourin, S. Baird, C.E. Neville, N. Wormskamp, B. Segers, M. Batzer, J. Lamerdin, P. de Jong, B. Wieringa, and R.G. Korneluk. 1993. Structure and genomic sequence of the myotonic dystrophy (DM kinase) gene. *Hum Mol Genet* **2**: 299-304.

Mahadevan, M., C. Tsilfidis, L. Sabourin, G. Shutler, C. Amemiya, G. Jansen, C. Neville, M. Narang, J. Barcelo, K. O'Hoy, S. Leblond, J. Earle-Macdonald, P.J. de Jong, B. Wieringa, and R.G. Korneluk. 1992. Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **255**: 1253-1255.

Mangiarini, L., K. Sathasivam, M. Seller, B. Cozens, A. Harper, C. Hetherington, M. Lawton, Y. Trottier, H. Lehrach, S.W. Davies, and G.P. Bates. 1996. Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* **87**: 493-506.

Mariappan, S.V.S., A.E. Garcia, and G. Gupta. 1996. Structure and dynamics of the DNA hairpins formed by tandemly repeated CTG triplets associated with myotonic dystrophy. *Nucleic Acids Res* **24**: 775-783.

Martorell, L., I. Illa, J. Rosell, J. Benitez, M.J. Sedano, and M. Baiget. 1996. Homozygous myotonic dystrophy: clinical and molecular studies of three unrelated cases. *J Med Genet* **33**: 783-785.

Martorell, L., J.M. Martinez, N. Carey, K. Johnson, and M. Baiget. 1995. Comparison of CTG repeat length expansion and clinical progression of myotonic dystrophy over a five year period. *J Med Genet* **32**: 593-596.

Matera, A.G., M.R. Frey, K. Margelot, and S.L. Wolin. 1995. A perinucleolar compartment contains several RNA polymerase III transcripts as well as the polypyrimidine tract-binding protein, hnRNP I. *J Cell Biol* **129**: 1181-1193.

Matunis, E.L., M.J. Matunis, and G. Dreyfuss. 1993. Association of individual hnRNP proteins and snRNPs with nascent transcripts. *J Cell Biol* **121**: 219-228.

Mayeda, A. and A.R. Krainer. 1992. Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* **68**: 365-375.

Mayrand, S., B. Setyono, J.R. Greenberg, and T. Pederson. 1981. Structure of nuclear ribonucleoprotein: identification of proteins in contact with poly(A)$^+$ heterogeneous nuclear RNA in living HeLa cells. *J Cell Biol* **90**: 380-384.

McKeown, M. 1992. Alternative mRNA splicing. *Ann Rev Cell Biol* **8**: 133-155.

Michael, W.M., M. Choi, and G. Dreyfuss. 1995. A nuclear export signal in hnRNP A1: a signal-mediated, temperature dependent nuclear protein export pathway. *Cell* **83**: 415-422.

Mitas, M. 1997. Trinucleotide repeats associated with human disease. *Nucleic Acids Res* **25**: 2245-2253.

Mitas, M., A. Yu, J. Dill, T.J. Kamp, E.J. Chambers, and I.S. Haworth. 1995. Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: $(CTG)_{15}$. *Nucleic Acids Res* **23**: 1050-1059.

Mizumoto, K. and Y. Kaziro. 1987. Messenger RNA capping enzymes from eukaryotic cells. *Prog Nucleic Acid Res Mol Biol* **34**: 1-28.

Monckton, D.G., T. Ashizawa, M.I. Coolbaugh, C.T. Caskey, M. Meistrich, and M.J. Siciliano. 1997a. Mouse Models of DM. In *First International Myotonic Dystrophy Consortium Conference*. Institut de Myologie/AFM, Paris.

Monckton, D.G., M.I. Coolbaugh, K.T. Ashizawa, M.J. Siciliano, and C.T. Caskey. 1997b. Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nat Genet* **15**: 193-196.

Monckton, D.G., L.J. Wong, T. Ashizawa, and C.T. Caskey. 1995. Somatic mosaicism, germline expansions, germline reversions and intergenerational reductions in myotonic dystrophy males: small pool PCR analyses. *Hum Mol Genet* **4**: 1-8.

Moore, P.B. 1996. The structure and function of 5S ribosomal RNA. In *Ribosomal RNA: structure, evolution, processing, and function in protein biosynthesis* (ed. R.A. Zimmerman and A.E. Dahlberg), pp. 199 - 236. CRC Press, Boca Raton.

Moreira, A., Y. Takagaki, S. Brackenridge, M. Wollerton, J.L. Manley, and N.J. Proudfoot. 1998. The upstream sequence element of the C2 complement poly(A) signal activates mRNA 3' end formation by two distinct mechanisms. *Genes Dev* **12**: 2522-2534.

Morrone, A., E. Pegoraro, C. Angelini, E. Zammarchi, G. Marconi, and E.P. Hoffman. 1997. RNA metabolism in myotonic dystrophy: patient muscle shows decreased insulin receptor RNA and protein consistent with abnormal insulin resistance. *J Clin Invest* **99**: 1691-1698.

Murphy, R. and S.R. Wente. 1996. An RNA-export mediator with an essential nuclear export signal. *Nature* **383**: 357-360.

Musco, G., G. Stier, C. Joseph, M.A. Castiglione Morelli, M. Nilges, T.J. Gibson, and A. Pastore. 1996. Three-dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell* **85**: 237-245.

Nagafuchi, S., H. Tanagisawa, K. Sato, T. Shirayama, and E. Ohsaki. 1994. Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat Genet* **6**: 14-18.

Nagai, K., C. Oubridge, T.H. Jessen, J. Li, and P.R. Eveans. 1990. Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature* **348**: 515-520.

Nakielny, S., U. Fischer, W.M. Michael, and G. Dreyfuss. 1997. RNA transport. *Ann Rev Neurosci* **20**: 269-301.

Napierala, M. and W.J. Krzyosiak. 1997. CUG repeats present in myotonin kinase RNA form metastable "slippery" hairpins. *J Biol Chem* **272**: 31079-31085.

Nayler, O., W. Stratling, J.P. Bourquin, I. Stagljar, L. Lindemann, H. Jasper, A.M. Hartmann, F.O. Fackelmayer, A. Ullrich, and S. Stamm. 1998. SAF-B protein couples transcription and pre-mRNA splicing to SAR/MAR elements. *Nucleic Acids Res* **26**: 3542-3549.

Neville, C.E., M.S. Mahadevan, J.M. Barcelo, and R.G. Korneluk. 1994. High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Hum Mol Genet* **3**: 45-51.

Nigg, E.A. 1997. Nucleocytoplasmic transport: signals, mechanisms and regulation. *Nature* **386**: 779-787.

Nothwang, H.G., T. Tamura, K. Tanaka, and A. Ichihara. 1994. Sequence analyses and inter-species comparisons of three novel human proteasomal subunits, HsN3, HsC7-I and HsC10-II, confine potential proteolytic active-site residues. *Biochim Biophys Acta* **1219**: 361-368.

Novelli, G., M. Gennarelli, G. Zelano, A. Pizzuti, C. Fattorini, C.T. Caskey, and B. Dallapiccola. 1993. Failure in detecting mRNA transcripts from the mutated allele in myotonic dystrophy muscle. *Biochem Mol Biol Int* **29**: 291-297.

Ohshima, K. and R.D. Wells. 1997. Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary desease genes. *J Biol Chem* **272**: 16798-16806.

Okano, H.J. and R.B. Darnell. 1997. A hierarchy of Hu RNA binding proteins in developing and adult neurons. *J Neurosci* **17**: 3024-3037.

Orr, H.T., M. Chung, S. Banfi, T.J. Kwiatkowski, and A. Servadio 1993. Expansion of an unstable trinucleotide CAG repeat in spino-cerebellar ataxia type 1. *Nat Genet* **4**: 221-226.

Otten, A.D. and S.J. Tapscott. 1995. Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci U S A* **92**: 5465-5469.

Paliouris, M. 1998. Role of Heterogeneous Nuclear Ribonucleoproteins in Myotonic Dystrophy. Master's Thesis. University of Florida, Gainesville.

Palliard, L., F. Omilli, V. Lagagneux, T. Bassez, D. Maniey, and H.B. Osborne. 1998. EDEN and EDEN-BP, a *cis* element and an associated factor that mediate sequence-specific mRNA deadenylation in *Xenopus* embryos. *EMBO J* **17**: 278-287.

Pasquinelli, A.E., R.K. Ernst, E. Lund, C. Grimm, M.L. Zapp, D. Rekosh, M.L. Hammarskjold, and J.E. Dahlberg. 1997. The constitutive transport element (CTE) of Mason-Pfizer monkey virus (MPMV) accesses a cellular mRNA export pathway. *EMBO J* **16**: 7500-7510.

Paulson, H.L. and K.H. Fischbeck. 1996. Trinucleotide repeats in neurogenetic disorders. *Ann Rev Neurosci* **19**: 79-107.

Peltomaki, P. 1997. DNA mismatch repair gene mutations in human cancer. *Environ Health Perspect* **105 Suppl 4**: 775-780.

Peng, S.S., C.Y. Chen, N. Xu, and A.B. Shyu. 1998. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *EMBO J* **17**: 3461-3470.

Petruska, J., N. Arnheim, and M.F. Goodman. 1996. Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases. *Nucleic Acids Res* **24**: 1992-1998.

Phillips, A.V., L.T. Timchenko, and T.A. Cooper. 1998. Disruption of splicing regulated by a CUG-binding protein in myotonic dystrophy. *Science* **280**: 737-741.

Piñol-Roma, S., Y.D. Choi, M.J. Matunis, and G. Dreyfuss. 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev* **2**: 215-227.

Piñol-Roma, S. and G. Dreyfuss. 1992. Shuttling of pre-mRNA binding proteins between nucleus and cytoplasm. *Nature* **355**: 730-732.

Pollard, V.W., W.M. Michael, S. Nakielny, M.C. Siomi, F. Wang, and G. Dreyfuss. 1996. A novel receptor-mediated nuclear protein import pathway. *Cell* **86**: 985-994.

Proudfoot, N. 1996. Ending the message is not so simple. *Cell* **87**: 779-781.

Ptacek, L.J., K.J. Johnson, and R.C. Griggs. 1993. Genetics and physiology of the myotonic muscle disorders. *New Engl J Med* **328**: 482-489.

Pulst, S.M., A. Nechiporuk, T. Nechiporuk, S. Gispert, X.N. Chen, I. Lopes-Cendes, S. Pearlman, S. Stakman, G. Orozco-Diaz, and A. Lunkes 1996. Moderate expansion

of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* **14**: 269-276.

Ranum, L.P., P.F. Rasmussen, K.A. Benzow, M.D. Koob, and J.W. Day. 1998. Genetic mapping of a second myotonic dystrophy locus. *Nat Genet* **19**: 196-198.

Reddy, P.S. and D.E. Housman. 1997. The complex pathology of trinucleotide repeats. *Curr Opin Cell Biol* **9**: 364-372.
Reddy, S., D.B. Smith, M.M. Rich, J.M. Leferovich, P. Reilly, B.M. Davis, K. Tran, H. Rayburn, R. Bronson, D. Cros, R.J. Balice-Gordon, and D. Housman. 1996. Mice lacking the myotonic dystrophy protein kinase develop a late onset progressive myopathy. *Nat Genet* **13**: 325-335.

Redman, J.B., R.G. Fenwick, Y.H. Fu, A. Pizzuti, and C.T. Caskey. 1993. Relationship between parental trinucleotide GCT repeat length and severity of myotonic dystrophy in offspring. *JAMA* **269**: 1960-1965.

Renz, A. and F.O. Fackelmayer. 1996. Purification and molecular cloning of the scaffold attachment factor B (SAF-B), a novel human nuclear protein that specifically binds to S/MAR- DNA. *Nucleic Acids Res* **24**: 843-849.

Richards, R.I. and G.R. Sutherland. 1994. Simple repeat DNA is not replicated simply. *Nat Genet* **6**: 114-116.

Richter, J.D. 1996. Dynamics of poly(A) addition and removal during development. In *Translational Control* (ed. J.W.B. Hershey, M.B. Mathews, and N. Sonenberg), pp. 481-503. Cold Spring Harbor Laboratory Press, Plainview, NY.

Ricker, K., M.C. Koch, F. Lehmann-Horn, D. Pongratz, N. Speich, K. Reiners, C. Schneider, and R.T. Moxley, 3rd. 1995. Proximal myotonic myopathy. Clinical features of a multisystem disorder similar to myotonic dystrophy. *Arch Neurol* **52**: 25-31.

Robberson, B.L., G.J. Cote, and S.M. Berget. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol Cell Biol* **10**: 84-94.

Roberts, R., N.A. Timchenko, J.W. Miller, S. Reddy, C.T. Caskey, M.S. Swanson, and L.T. Timchenko. 1997. Altered phosphorylation and intracellular distributionof a $(CUG)_n$ triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice. *Proc Natl Acad Sci U S A* **94**: 13221-13226.

Robinow, S. and K. White. 1988. The locus elav of *Drosophila melanogaster* is expressed in neurons at all developmental stages. *Dev Biol* **126**: 294-303.

Roses, A.D. and S.H. Appel. 1973. Protein kinase activity in erythrocyte ghosts of patients with myotonic muscular dystrophy. *Proc Natl Acad Sci U S A* **70**: 1855-1859.

Roses, A.D. and S.H. Appel. 1974. Muscle membrane protein kinase in myotonic muscular dystrophy. *Nature* **250**: 245-247.

Ross, J. 1995. mRNA stability in mammalian cells. *Microbiol Rev* **59**: 423-450.

Ryan, K.J. and T.A. Cooper. 1996. Muscle-specific splicing enhancers regulate inclusion of the cardiac troponin T alternative exon in embryonic skeletal muscle. *Mol Cell Biol* **16**: 4014-4023.

Sabourin, L.A., M.S. Mahadevan, M. Narang, D.S.C. Lee, L.C. Surh, and R.G. Korneluk. 1993. Effect of the myotonic dystrophy (DM) mutation mRNA levels of the DM gene. *Nat Genet* **4**: 233-238.

Salditt-Georgieff, M., M. Harpold, S. Chen-Kiang, and J.E. Darnell, Jr. 1980. The addition of 5' cap structures occurs early in hnRNA synthesis and prematurely terminated molecules are capped. *Cell* **19**: 69-78.

Sambrook, J., E. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sanpei, K., H. Takano, S. Igarashi, T. Sato, M. Oyake, H. Sasaki, A. Wakisaka, K. Tashiro, Y. Ishida, and T. Ikeuchi 1996. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* **14**: 277-284.

Scherly, D., W. Boelens, N.A. Dathan, W.J. van Venrooij, and I.W. Mattaj. 1990. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B" and their cognate RNAs. *Nature* **345**: 502-506.

Shatkin, A.J. 1985. mRNA cap binding proteins: essential factors for initiating translation. *Cell* **40**: 223-224.

Shaw, D.J., M. McCurrach, S.A. Rundle, H.G. Harley, S.R. Crow, R. Sohn, J.P. Thirion, M.G. Hamshere, A.J. Buckler, and P.S. Harper. 1993. Genomic organization and transcriptional units at the myotonic dystrophy locus. *Genomics* **18**: 673-679.

Sheets, M.D. and M. Wickens. 1989. Two phases in the addition of a poly(A) tail. *Genes Dev* **3**: 1401-1412.

Siomi, H. and G. Dreyfuss. 1995. A nuclear localizaion domain in the hnRNP A1 protein. *J Cell Biol* **129**: 551-560.

Siomi, H. and G. Dreyfuss. 1997. RNA-binding proteins as regulators of gene expression. *Curr Opin Genet Dev* **7**: 345-353.

Siomi, M.C., M. Fromont, J.C. Rain, L. Wan, F. Wang, P. Legrain, and G. Dreyfuss. 1998. Functional conservation of the transportin nuclear import pathway in divergent organisms. *Mol Cell Biol* **18**: 4141-4148.

Siomi, H., M.J. Matunis, W.M. Michael, and G. Dreyfuss. 1993. The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nucleic Acids Res* **21**: 1193-1198.

Siomi, M.C., Y. Zhang, H. Siomi, and G. Dreyfuss. 1996. Specific sequences in the fragile X syndrome protein FMR1 and the FXR proteins mediate their binding to 60S ribosomal subunits and the interactions among them. *Mol Cell Biol* **16**: 3825-3832.

Smith, G.K., J. Jie, G.E. Fox, and X. Gao. 1995. DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes. *Nucleic Acids Res* **23**: 4303-4311.

St. Johnston, D. 1995. The intracellular localization of messenger RNAs. *Cell* **81**: 161-170.

Stade, K., C.S. Ford, C. Guthrie, and K. Weis. 1997. Exportin 1 (Crm1p) is an essential nuclear export factor. *Cell* **90**: 1041-1050.

Steinback, P., D. Glaser, W. Vogel, M. Wolf, and S. Schwemmle. 1998. The DMPK gene of severely affected myotonic dystrophy patients is hypermethylated proximal to the largely expanded CTG repeat. *Am J Hum Genet* **62**: 278 - 285.

Stutz, F., M. Neville, and M. Rosbash. 1995. Identification of a novel nuclear pore-associated protein as a functional target of the HIV-1 Rev protein in yeast. *Cell* **82**: 495-506.

Swanson, M.S. 1995. Functions of nuclear pre-mRNA/mRNA binding proteins. In *Pre-mRNA processing* (ed. A.I. Lamond). R.G. Landes Company.

Swanson, M.S. and G. Dreyfuss. 1988a. RNA binding specificity of hnRNP proteins: a subset bind to the 3' end of introns. *EMBO J* **7**: 3519-3529.

Swanson, M.S. and G. Dreyfuss. 1988b. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol Cell Biol* **8**: 2237-2241.

Swanson, M.S., T.Y. Nakagawa, K. LeVan, and G. Dreyfuss. 1987. Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and

domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. *Mol Cell Biol* **7**: 1731-1739.

Takagaki, Y., R.L. Seipelt, M.L. Peterson, and J.L. Manley. 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. *Cell* **87**: 941-952.

Taneja, K.L., M. McCurrach, M. Schalling, D. Housman, and R.H. Singer. 1995. Foci of trinucleotide repeat transcripts in nuclei of myotonic dystrophy cells and tissues. *J Cell Biol* **128**: 995-1002.

Thornton, C., R.C. Griggs, and R.T. Moxley, 3rd. 1995. PROMM syndrome (Ricker's disease). *Ann Neurol* **38**: 273.

Thornton, C.A., J.P. Wymer, Z. Simmons, C. McClain, and R.T. Moxley, 3rd. 1997. Expansion of the myotonic dystrophy CTG repeat reduces expression of the flanking DMAHP gene. *Nat Genet* **16**: 407-409.

Timchenko, L.T., J.W. Miller, N.A. Timchenko, D.R. DeVore, K.V. Datar, L. Lin, R. Roberts, C.T. Caskey, and M.S. Swanson. 1996a. Identification of a $(CUG)_n$ triplet repeat RNA-binding protein and its expression in myotonic dystrophy. *Nucleic Acids Res* **24**: 4407-4414.

Timchenko, L., W. Nastainczyk, T. Schneider, B. Patel, F. Hofmann, and C.T. Caskey. 1995. Full-length myotonin protein kinase (72 kDa) displays serine kinase activity. *Proc Natl Acad Sci U S A* **92**: 5366-5370.

Timchenko, L.T., N.A. Timchenko, C.T. Caskey, and R. Roberts. 1996b. Novel proteins with binding specificity for DNA CTG repeats and RNA CUG repeats: implications for myotonic dystrophy. *Hum Mol Genet* **5**: 115-121.

Timchenko, N., D.R. Wilson, L.R. Taylor, S. Abdelsayed, M. Wilde, M. Sawadogo, and G.J. Darlington. 1995. Autoregulation of the human C/EBP alpha gene by stimulation of upstream stimulatory factor binding. *Mol Cell Biol* **15**: 1192-1202.

Tishkoff, D.X., N. Filosi, G.M. Gaida, and R.D. Kolodner. 1997. A novel mutation avoidance mechanism dependent on S. cerevisiae RAD27 is distinct from DNA mismatch repair. *Cell* **88**: 253-263.

Todorov, I.T., A. Attaran, and S.E. Kearsey. 1995. BM28, a human member of the MCM2-3-5 family, is displaced from chromatin during DNA replication. *J Cell Biol* **129**: 1433-1445.

Trottier, Y., Y. Lutz, G. Stevanin, G. Imbert, D. Devys, G. Cancel, F. Saudou, C. Weber, G. David, and L. Tora 1995. Polyglutamine expansion as a pathologica

epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* **378**: 403-406.

Truant, R., R.A. Fridell, R.E. Benson, H. Bogerd, and B.R. Cullen. 1998. Identification and functional characterization of a novel nuclear localization signal present in the yeast Nab2 poly(A)+ RNA binding protein. *Mol Cell Biol* **18**: 1449-1458.

van der Ven, P.F., G. Jansen, T.H. van Kuppevelt, M.B. Perryman, M. Lupa, P.W. Dunne, H.J. ter Laak, P.H. Jap, J.H. Veerkamp, H.F. Epstein, and B. Wieringa. 1993. Myotonic dystrophy kinase is a component of neuromuscular junctions. *Hum Mol Genet* **2**: 1889-1894.

Verkerk, A.J., M. Pieretti, J.S. Sutcliffe, Y.H. Fu, and D.P. Kuhl 1991. Identification of a gene (FMR-1) containing a CGG repeat coincidint with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**: 905-914.

Visa, N., A.T. Alzhanova-Ericsson, X. Sun, E. Kiseleva, B. Bjorkroth, T. Wurtz, and B. Daneholt. 1996. A pre-mRNA-binding protein accompanies the RNA from the gene througgh the nuclear pores and into polysomes. *Cell* **84**: 253-264.

Wahle, E. 1991. A novel poly(A)-binding protein acts as a specificity factor in the second phase of messenger RNA polyadenylation. *Cell* **66**: 759-768.

Wahle, E. 1995. Poly(A) tail length control is caused by termination of processive synthesis. *J Biol Chem* **270**: 2800-2808.

Wahle, E. and W. Keller. 1996. The biochemistry of polyadenylation. *Trends Biochem Sci* **21**: 247-250.

Wahle, E. and U. Kuhn. 1997. The mechanism of 3' cleavage and polyadenylation of eukaryotic pre-mRNA. *Prog Nucleic Acid Res Mol Biol* **57**: 41-71.

Wakamatsu, Y. and J.A. Weston. 1997. Sequential expression and role of Hu RNA-binding proteins during neurogenesis. *Development* **124**: 3449-3460.

Wang, J., E. Pegoraro, E. Menegazzo, M. Gennarelli, R.C. Hoop, C. Angelini, and E.P. Hoffman. 1995. Myotonic dystrophy: evidence for a possible dominant-negative RNA mutation. *Hum Mol Genet* **4**: 599-606.

Wang, Y.H., S. Amirhaeri, S. Kang, R.D. Wells, and J.D. Griffith. 1994. Preferential nucleosome assembly at DNA triplet repeats from the myotonic dystrophy gene. *Science* **265**: 669-671.

Wang, Y.H. and J. Griffith. 1995. Expanded CTG triplet blocks from the myotonic dystrophy gene create the strongest known natural nucleosome positioning elements. *Genomics* **25**: 570-573.

Wang, Y.X., S. Huang, and D.E. Draper. 1996. Structure of a U.U pair within a conserved ribosomal RNA hairpin. *Nucleic Acids Res* **24**: 2666-2672.

Waring, J.D., R. Haq, K. Tamai, L.A. Sabourin, J.E. Ikeda, and R.G. Korneluk. 1996. Investigation of myotonic dystrophy kinase isoform translocation and membrane association. *J Biol Chem* **271**: 15187-15193.

Warrick, J.M., H.L. Paulson, G.L. Gray-Board, Q.T. Bui, K.H. Fischbeck, R.N. Pittman, and N.M. Bonini. 1998. Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in Drosophila. *Cell* **93**: 939-949.

Wells, R.D. 1996. Molecular basis of genetic instability of triplet repeats. *J Biol Chem* **271**: 2875-2878.

Whiting, E.J., J.D. Waring, K. Tamai, M.J. Somerville, M. Hincke, W.A. Staines, J.E. Ikeda, and R.G. Korneluk. 1995. Characterization of myotonic dystrophy kinase (DMK) protein in human and rodent muscle and central nervous tissue. *Hum Mol Genet* **4**: 1063-1072.

Wilson, S.M., K.V. Datar, M.R. Paddy, J.R. Swedlow, and M.S. Swanson. 1994. Characterization of nuclear polyadenylated RNA-binding proteins in Saccharomyces cerevisiae. *J Cell Biol* **127**: 1173-1184.

Wong, L.-J.C., T. Ashizawa, E.G. Monckton, C.T. Caskey, and C.S. Richards. 1995. Somatic heterogeneity of the CTG repeat in myotonic dystrophy is age and size dependent. *Am J Hum Genet* **56**: 114-122.

Wooster, R., A.M. Cleton-Jansen, N. Collins, J. Mangion, R.S. Cornelis, C.S. Cooper, B.A. Gusterson, B.A.J. Ponder, A. von Deimling, O.D. Wiestler, C.J. Cornelisse, P. Devilee, and M.R. Stratton. 1994. Instability of short tandem repeats (microsatellites) in human cancers. *Nat Genet* **6**: 152-156.

Yu, S., M. Mangelsdorf, D. Hewett, L. Hobson, E. Baker, H.J. Eyre, N. Lapsys, D. Le Paslier, N.A. Doggett, G.R. Sutherland, and R.I. Richards. 1997. Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. *Cell* **88**: 367-374.

Zhang, W., B.J. Wagner, K. Ehrenman, A.W. Schaefer, C.T. DeMaria, D. Crater, K. DeHaven, L. Long, and G. Brewer. 1993. Purification, characterization, and cDNA cloning of an AU-rich element RNA-binding protein, AUF1. *Mol Cell Biol* **13**: 7652-7665.

Zhuchenko, O., J. Bailey, P. Bonnen, T. Ashizawa, D.W. Stockton, C. Amos, W.B. Dobyns, S.H. Subramony, H.Y. Zoghbi, and C.C. Lee. 1997. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat Genet* **15**: 62-69.

BIOGRAPHICAL SKETCH

Jill W. Miller was born and raised in Melbourne, Florida. She is the younger of two daughters born to John W. Watkins and Robbie C. Watkins. She started college at the University of Florida in 1987 and obtained a Bachelor of Science degree in Biochemistry in 1991. She attended medical school at the University of Florida College of Medicine and is currently enrolled in a combined M.D./Ph.D. program. She began graduate school in the summer of 1994 at the University of Florida in the Department of Molecular Genetics and Microbiology. She completed her thesis under the guidance of Dr. Maurice Swanson, and received her Doctorate in 1998. She will be continuing on with her third year of medical school this fall.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Maurice S. Swanson, Chair
Associate Professor of Molecular
Genetics and Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

John P. Aris
Assistant Professor of Anatomy
and Cell Biology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Henry V. Baker
Associate Professor of Molecular
Genetics and Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

James B. Flanegan
Professor of Molecular
Genetics and Microbiology

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Gerard P. Shaw
Professor of Neuroscience

This dissertation was submitted to the Graduate Faculty of the College of Medicine and to the Graduate School and was accepted as partial fulfillment for the degree of Doctor of Philosophy.

December, 1998

Dean, College of Medicine

Dean, Graduate School